Hum Genet (2008) 123:177–187
DOI 10.1007/s00439-007-0460-x

ORIGINAL INVESTIGATION

Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression

Hans Eiberg · Jesper Troelsen · Mette Nielsen · Annemette Mikkelsen · Jonas Mengel-From · Klaus W. Kjaer · Lars Hansen

Received: 31 October 2007 / Accepted: 18 December 2007 / Published online: 3 January 2008

Abstract The human eye color is a quantitative trait displaying multifactorial inheritance. Several studies have shown that the *OCA2* locus is the major contributor to the human eye color variation. By linkage analysis of a large Danish family, we finemapped the blue eye color locus to a 166 Kbp region within the *HERC2* gene. By association analyses, we identified two SNPs within this region that were perfectly associated with the blue and brown eye colors: rs12913832 and rs1129038. Of these, rs12913832 is located 21.152 bp upstream from the *OCA2* promoter in a highly conserved sequence in intron 86 of *HERC2*. The brown eye color allele of rs12913832 is highly conserved throughout a number of species. As shown by a Luciferase assays in cell cultures, the element significantly reduces the activity of the *OCA2* promoter and electrophoretic mobility shift assays demonstrate that the two alleles bind different subsets of nuclear extracts. One single haplotype, represented by six polymorphic SNPs covering half of the 3′ end of the *HERC2* gene, was found in 155 blue-eyed individuals from Denmark, and in 5 and 2 blue-eyed individuals from Turkey and Jordan, respectively. Hence, our data suggest a common founder mutation in an *OCA2* inhibiting regulatory element as the cause of blue eye color in humans. In addition, an LOD score of $Z = 4.21$ between hair color and D14S72 was obtained in the large family, indicating that *RABGGTA* is a candidate gene for hair color.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-007-0460-x) contains supplementary material, which is available to authorized users.

H. Eiberg (✉) · J. Troelsen · M. Nielsen · A. Mikkelsen · K. W. Kjaer · L. Hansen
Department of Cellular and Molecular Medicine, Section IV Build. 24.4, Panum Institute, University of Copenhagen, Blegdamsvej 3b, 2200 Copenhagen, Denmark
e-mail: he@imbg.ku.dk

J. Mengel-From
Institute of Forensic Genetics, University of Copenhagen, Fredrik V's vej 11, 2100 Copenhagen, Denmark

K. W. Kjaer · L. Hansen
The Wilhelm Johannsen Centre for Functional Genome Research, Department of Cellular and Molecular Medicine, Panum Institute, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen, Denmark

Introduction

The eye color variation is a result of variable amounts and distribution of melanin pigment in the iris. In humans, the most remarkable eye color diversity is found among Caucasians. Blue/brown eye color genetics are known to the public as school example of monogenic inheritance, however, the variation in pigment concentration and the distribution of pigment in the iris (Fig. 1) suggest the eye color genetics to be far more complex as supported by recent data (Eiberg and Mohr 1996; Sturm et al. 2001; Zhu et al. 2004; Posthuma et al. 2006; Duffy et al. 2007). Loss of pigmentation in the skin, hair and eyes, also known as oculocutaneous albinism, derives from mutations in the genes *OCA1-4* (Sturm et al. 2001). The locus responsible for the brown or blue eye color phenotypes (MIM 227220) was first identified by linkage to chromosome 15q (multipoint LOD score of $Z = 32.2$) in the Danish population by Eiberg and Mohr (1996). The candidate locus was mapped to 15q12-13 flanked by the markers D15S156 and D15S144 with a maximum LOD score close to D15S165 and *OCA2* (MIM 611409) was suggested as the candidate gene. Subse-

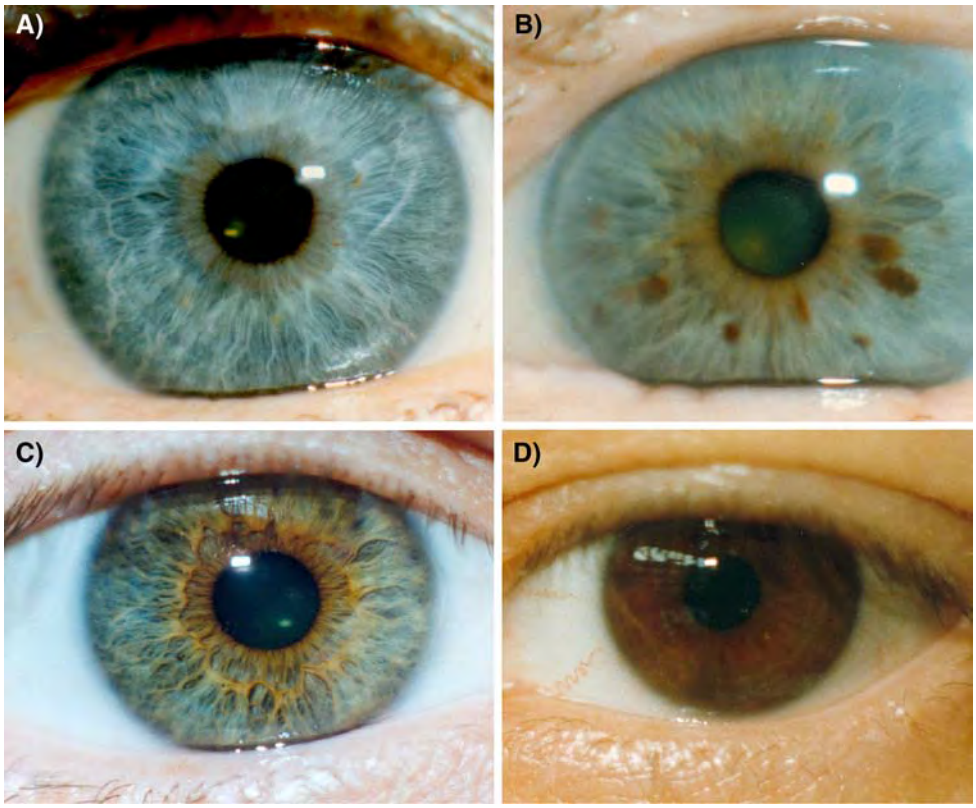

Fig. 1 Different eye color phenotypes. **a** *Blue* (without *brown* areas). **b** *Blue* with *brown spots* (with *brown*) scored as "unknown" in the linkage and association studies. **c** *Brown–green/hazel* (BEY1) with a board puripupillary ring. **d** *Brown* (BEY2), total *brown* pigmentation. The person with *blue* eye color (**a**) represents the genotype rs12913832 G/G, while the persons **b**–**d** represent the genotype rs12913832 A/G

quently, several other linkage and association studies confirmed the locus and *OCA2* as the major contributors to human eye color variation (Sturm et al. 2001; Zhu et al. 2004; Posthuma et al. 2006; Frudakis et al. 2003) and 74% of the variation was estimated to a QTL, linked to *OCA2* (Duffy et al. 2007).

Attempts to identify the BEY/blue mutations in *OCA2* have failed (Frudakis et al. 2007), and a variation within the 5′ proximal regulatory control region of *OCA2* was suggested to be responsible for 90% of the eye color pigmentation variation (Duffy et al. 2007).

The function of the *OCA2* protein is ambiguous (Rebbeck et al. 2002), and it has been suggested to be a Na^+/H^+ antiporter (Ancans et al. 2001; Puri et al. 2000) or a glutamate transporter (Lamoreux et al. 1995). Both the functions indicate that *OCA2* is involved in the supply of substrates to a tyrosinase in the biosynthesis of melanin (Ancans et al. 2001). Alternatively, *OCA2* may be involved in the intracellular trafficking of the tyrosinase enzyme during melanosome maturation (Toyofuku et al. 2002).

Upstream of *OCA2* is the *HERC2* gene (MIM 605837) located. Deletions of exon 86–90 or exon 53–86 in *Herc2* of mice have been shown to affect eye and coat color pigmentation as well as sperm production leading to male sterility (Lehman et al. 1998; Russel et al. 1995). In humans, partial deletions of the *OCA2-HERC2* locus are known in the Prader-Willi and Angelman syndromes. Both syndromes have been associated with oculocutaneous albinism or reduced pigmentation (Spritz et al. 1997) in several cases. The function of *HERC2* is unknown but the gene encodes for deferent conserved functional protein domains involved in spermatogenesis, ubiqutin mediated proteolysis and intracellular transport (Yonggang et al. 2000). The human genome encodes for 3–12 *HERC2* pseudo genes, most located on chromosome 15q, which have complicated studies of this region.

Here we present evidence from linkage and association studies that a region in *HERC2* contains a highly conserved regulatory element, which is the cause of blue eye color in humans. The influence of other loci like hair color, as controlling of the variation in the brown eye colors, are examined.

Material and methods

Family material

A three-generation Danish family (CFB#694) representing 28 informative meioses was used for linkage analysis and the blue eye color locus was finemapped. The family used in the linkage analysis, association and haplotype studies were of Danish origin and retrieved from the Copenhagen Family Bank, the families used in this study (families

CFB#604-1505) (Eiberg et al. 1989). Only families with siblings, who had blue and brown eyes, respectively, were included in the study. The parents and siblings were classified as blue-eyed (Fig. 1a) or brown-eyed (Fig. 1c or d) individuals. Haplotypes were constructed from 100 Danish informative selected trios families, and most of these parents were also included in the association studies. These 100 triosis represented 45 families where at least one individual had brown eyes and 55 families where all individuals had blue eyes. Families where green and brown eye color spots segregate were not used. The haplotypes were deduced manually from the family study. Additional control material for DNA sequencing was collected from two large Danish families from the Copenhagen Family Bank. Five individuals from Turkey with blue eyes, black hair and light skin and two individual from Jordan with blue eyes, black hair and dark skin were included in the association analysis. Additionally, two persons with natal heterochromia were examined.

The blue eye color phenotype was defined as a complete lack of brown pigmentation (Fig. 1a), an intermediate phenotype was defined as "blue eye with blown dots" (Fig. 1b), an intermediate brown eye color phenotype was defined as hazel with a broad peripupillary ring and was named the BEY1 phenotype (Fig. 1c), and a complete brown pigmented eye color was defined as the BEY2 phenotype (Fig. 1d).

All individuals in the study were interviewed by questionnaires and asked to determine their own eye color from the categories: brown, blue, gray and green, and whether brown spots or brown peripupillar rings were present.

Hair colors were categorized as red, black, brown and blond hair at the time when the persons were between 20 and 30 years of age. In family CFB#694, the eye color for all individuals was documented by photos and all key persons were re-examined. All individuals with green eye color or blue or gray eye color with brown spots not located close to the pupil were excluded from the linkage and association studies. Genomic DNA was extracted from the whole blood using standard phenol/chloroform procedures and the study adhered to the tenets of the declaration of Helsinki.

Association studies

Database search revealed that the majority of the SNPs in *HERC2* display low frequencies [$q < 0.1$, UCSC Human Genome Browser (July 2003) and GenBank dbSNP] and were not included in the association study due to lack of potent information. Many of the reported ~2.000 SNPs in *HERC2* can be explained by single base pair variations between *HERC2* and the *HERC2* pseudo genes. After exclusion of redundant SNPs, we were able to identify 15 polymorphic SNPs (supplementary Table 1) which were used for association analysis and construction of haplotypes (Tables 1, 2). Fishers exact test was used to test for significant genotype–phenotype distributions.

Linkage analysis and haplotype construction

The candidate region *OCA2-HERC2-APBA2* for the BEY1 locus (Eiberg and Mohr 1996) was finemapped and haplotypes were constructed using the following STS markers and SNPs: D15S1002, D15S156, D15S1533, rs4074658, rs1129038, rs10680280, rs3054537, rs10627597 and D15S1048 (Fig. 2). SNPs were analyzed by direct DNA sequencing or restriction enzyme digests (PCR primer sequences are available in supplementary Table 1). Additional markers were tested and found non-informative in the analyzed families (data not shown). LOD score calculations were done using FASTLINK (Schäffer et al. 1994) and brown eye color was considered as a dominant trait to the blue eye color. The CFG#694 family members were previously genotyped with more than 400 markers (Eiberg and Mohr 1996).

DNA sequencing

Four unrelated individuals representing the blue eye color phenotype (Fig. 1a), the BEY1 variant phenotype with several brown spots (Fig. 1b), the BEY1 (Fig. 1c) and the BEY2 phenotype, respectively, (Fig. 1c, d) were selected for identification of polymorphic nucleotide positions in the BEY candidate region by DNA sequence analysis. The three persons with the different brown eye color phenotypes were selected from families, where LOD scores higher than $Z = 2.5$ for the markers D15S1533 and rs10680280 in the BEY candidate region were achieved. The *OCA2* promoter (Lee et al. 1995) and exon 1, 2, 7 and 10 and *HERC2* exon 21, 26, 30, 33, 39, 44–47, 51, 52, 54, 55, 59, 72, 76–78, 85, 87–93 including the 3′UTR sequence were sequenced bidirectionally.

The sequenced regions represented coding exons including non-synonymous SNPs, CpG islands, and regions coding for zink finger domains or phylogenetic conserved areas located in introns of *OCA2* and *HERC2* according to the UCSC-genome browser (primer positions and physical positions are given in supplementary Table 2). Oligonucleotides were designed using the software Primer3 in combination with ClustalW alignments for identification of unique primer sequences. Exon PCR amplification included minimum 100 bp of the intron–exon splice site. The primers used for sequence analyses of polymorphic positions in *HERC2* are presented in supplementary Table 2.

PCR was carried out using standard conditions according to the manufactures protocols, Taq DNA polymerases

Table 1 Association of 12 SNPs in the *OCA2-HERC2-APBA2* locus

OCA2	rs4778190 (int24)			rs1800401 (ex9)			rs4778241 (int1)		
	AA	AT	TT	GG	GA	AA	CC	CA	AA
Without brown/blue	10	6	3	91	6	0	59	3	0
With brown	15	13	2	35	4	0	15	30	0
	P = 0.48			P = 0.47			P = 3.13e–12		
HERC2	rs1129038 (3′UTR)			rs12913832 (int86)			rs3935591 (int83)		
	AA	AG	GG	GG	AG	AA	GG	AG	AA
Without brown/blue	150	0	0	150	0	0	150	0	0
With brown	0	46	0	0	46	0	15	31	0
	P = 6.12e–46			P = 6.12e–46			P = 1.5e–25		
HERC2	rs11636232 (ex78)			rs7170852 (int56)			rs2238289 (int44)		
	CC	TC	TT	AA	AT	TT	TT	TC	CC
Without brown/blue	12	20	8	147	3	0	150	0	0
With brown	15	14	0	19	27	0	18	28	0
	P = 0.012			P = 1.1e–17			P = 4.2e–22		
HERC2/APBA2	rs2240203 (int20)			rs916977 (int12)			rs2279483 (APBA2,int10)		
	AA	AG	GG	GG	AG	AA	AA	AG	GG
Without brown/ blue	149	1	0	145	5	0	55	18	2
With brown	36	29	0	15	30	0	28	4	0
	P = 8.9e–17			P = 3.9e–19			P = 0.31		

The DNA samples represent 150 blue eyed and 46 brown-eyed Danish parents retrieved from Copenhagen Family Bank (CFB). *P* values were calculated using Fisher exact test

Table 2 Haplotypes for 13 different SNPs in *OCA2* and *HERC2* found by examination of 100 family triads

Phenotype	Blue				Brown					
Haplotype[a]	h-1	h-2	h-3	h-4	h-5	h-6	h-7	h-8	h-9	h-10
rs4778241-OCA2[b]	C (*A*[3])	C	C	C	C (*A*[2])	*A*	*A*	*A* (C[1])	*A*	*A*
rs1129038[c]	A	q	A	A	*G*	*G*	*G*	*G*	*G*	*G*
rs12593929	A	A	A	A	A	A	A	A	*G*	*G*
rs12913832[c]	G	G	G	G	*A*	*A*	*A*	*A*	*A*	*A*
rs7183877	C	C	C	C	C	C	C	*A*	*A*	*A*
rs3935591	G	G	G	G	G	*A*	*A*	*A*	*A*	*A*
rs7170852	A	A	*T*	A	A	A	A	*T*	A	*T*
rs2238289	T	T	T	T	T	T	T	*C*	*C*	*C*
rs3940272	C	C	C	C	C	C	C	*A*	*A*	*A*
rs8028689	T	T	T	T	T	T	T	T	*C*	*C*
rs2240203	A	A	A	*G*	A	A	A	A	*G*	G
rs11631797	G	G	G	*A*	G	G	G	*A*	*A*	*A*
rs916977	G	*A*	G	*A*	G	G	*A*	*A*	*A*	*A*
Number	345	5	4	1	14	1	3	19	1	7
Total number	355				45					

Only the father and the mother were analyzed in the 100 families. The families were retrieved from the Copenhagen Family Bank[1]

[a] Haplotype #1 to #4 represent individuals with blue eye color; haplotype #5 to #10 represent individuals with brown eye color (BEY1 or BEY2)

[b] Values in italics (alleles and number in the brackets) represent possible new mutations or recombinants

[c] Variations in rs1129038 and rs12913832 are common in Caucasians and very rare among other ethnic groups (dbSNP, GenBank, NCBI)

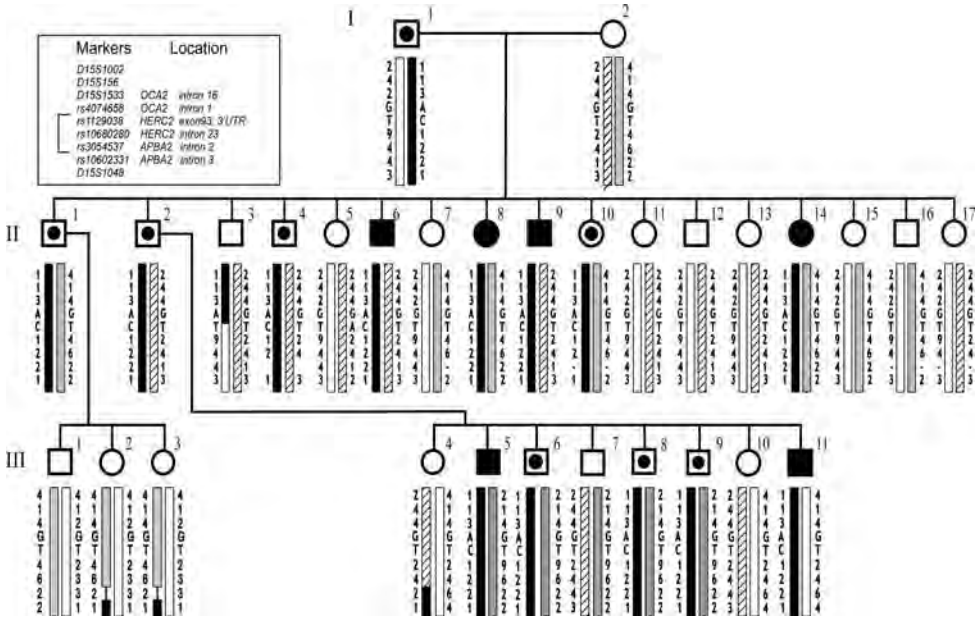

Fig. 2 Three generation's pedigree for the family 694 segregating for the brown eye color. The iris phenotypes are defined as blue (symbols *open square, open circle*), BEY1 is a broad brown area around the pupil (peripupillary ring, symbols *filled square, filled circle*) and complete brown eye color BEY2 (symbols *filled square, filled circle*). The BEY2 phenotype was darker in a peripupillary ring

were purchased from Qiagen (Hilden, Germany), New England Biolabs (Ipswich, MA, USA.) and Applied Biosystem Industries (Foster City, CA, USA.). Reactions were carried out in 15 µl volumes containing the buffer, 2.5 µM dNTP, 10 µM of each primer, 0.008% cresol red (Sigma–Aldrich Co., St Louis, USA), 12% sucrose (w/v), and 50–100 ng templates DNA. Standard reaction conditions for all primer pairs were: 95°C, 5 min; 40 cycles 95°C for 30 s, 56.4°C for 30 s and 72°C for 1 min; followed by 5 min at 72°C. PCR reactions were analyzed by 2% agarose gel-electrophoresis stained with ethidiumbromide, 1× TBE, before sequencing. Sequencing was carried out according to the manufactures protocol using the PCR primers and BigDye ver 1.1 (Applied Biosystems, Foster City, CA, USA) without further modifications and analyzed using an ABI370 sequencer. Sequence data were analyzed using standard software (Chromas ver. 2.1, Technelysium Pty Ltd, Australia) and sequence alignments were carried out using ClustalW. Restriction enzyme digests of PCR products were separated by 2% agarose gel-electrophoresis, 1× TBE.

Promoter activity assay in cell culture

The human *OCA2* promoter region position +80 to −435 was PCR amplified using genomic DNA from a brown-eyed homozygous individual using the *Hin*dIII and *Kpn*I primer pair (supplementary Table 2). The PCR product was subcloned into pCR2.1-TOPO and confirmed by sequencing. The *Hin*dIII/*Kpn*I *OCA2* promoter fragment was further subcloned into the luciferase reporter plasmid pGL4.10 generating the plasmid pGL4 OCA2. The region in *HERC2* intron 86 located at position −20,708 to −21,383 upstream from the *OCA2* transcriptional start site was PCR amplified using the *Bam*HI containing primer pair and genomic DNA from one blue-eyed person (the G-allele of rs12913832) and from one brown eyed (BEY2) person (the A-allele of rs12913832) (supplementary Table 2). The 675 bp PCR fragments were cloned into pCR2.1-TOPO and sequenced before subcloning into pGL4 *OCA2* using the *Bam*HI restriction site upstream of the *OCA2* promoter. The two plasmids were named pGL4 *OCA2-HERC2*-blue and pGL4 *OCA2-HERC2*-brown. The transcriptional activity of the plasmid pGL4 *OCA2*, pGL4 *OCA2-HERC2*-blue and pGL4 *OCA2-HERC2*-brown were analyzed by transfection into Caco-2 and COS7 cells (Remenyi et al. 2004; Troelsen et al. 2003b), and the luciferase activities were corrected for transfection efficiency and normalized to the level of pGL4 OCA2.

Preparation of nuclear extracts and EMSA were performed as previously described (Troelsen et al. 2003a). The following oligonucleotides are used in the assay:

A-variant: tcatttgagcattaaatgtcaagttctgcacgc and agcgtg cagaacttgacatttaatgctcaaatg;

G-variant: tcatttgagcattaagtgtcaagttctgcacgc and agcgtg cagaacttgacacttaatgctcaaatg

Unspec: aacgtagctgatcgaatcggttac and agtaaccgattcg atcagctacgt

Results

The brown (BEY2) and the hazel (BEY1) eye color phenotypes segregate in the Danish family CFB#694 (Fig. 2). Linkage analysis resulted in an LOD score of $Z = 6.30$ ($\theta = 0$) for marker rs10680280 when segregation of the BEY2 phenotype was considered and an LOD score of

$Z = 8.10$ ($\theta = 0$) when the BEY2 and the BEY1 phenotypes were considered as one phenotype. Haplotypes were constructed for five microsatellites and four SNPs and revealed two key recombination events in individuals II-3 and III-4, respectively. This narrowed down the candidate region to be between rs4074658 and rs10602331 (Figs. 2, 3a).

Association studies

The SNP allele distribution in the candidate region was studied in 155 blue-eyed and 45 brown-eyed unrelated individuals (Table 1). A complete association was found for the alleles rs1129038*A and rs12913832*G (*P*-value 6.12 e-46) with blue eye phenotype and these two candidate SNP alleles were found to be in c*is* position (Table 1).

The haplotype study

Haplotypes using 13 SNPs in the *OCA2-HERC2* locus were associated with brown and blue eye color in 200 parents from 100 trios. Ten different haplotypes were found where the haplotypes h-1 to h-4 represented the blue eye color and the haplotypes h-5 to h-10 represented the brown eye color phenotypes BEY1 and BEY2 (Table 2).

One recombination event was observed between rs8028689 and rs2240203 in a family carrying haplotype h-4. This recombination could minimize the BEY candidate region to a 166 Kbp fragment between the SNPs rs4074658 (Fig. 3a) and rs2240203, which covers the region from *OCA2* intron 1 to *HERC2* intron 20 (Table 2; Fig. 3b).

The rs4778241*A allele for haplotype 1 found in three persons could be a later mutation or a recombination event, since this marker is close to rs4074658. We have not examined this recombination further. The haplotype h-1 was found as the common haplotype among blue-eyed persons from Denmark and the haplotype was further found in seven unrelated individuals with the blue eye phenotype from Turkey or Jordan and in two persons with natal heterochromia.

DNA sequencing

The following *HERC2* exons 21, 26, 30, 33, 39, 44–47, 51, 52, 54, 55, 59, 72, 76–78, 85, 87–93 including 3′UTR and the *OCA2* exon 1, 2, 7 and 10 and the *OCA2* promoter were sequenced without detecting any causative DNA variations in the coding regions, the intron-splicing regions and the *OCA2* promoter region. Sequence analysis of the corresponding regions in genomic DNA from the two persons with heterochromia showed that both the persons carried the blue h-1 haplotype.

Functional analysis of rs12913832

To elucidate possible regulatory effects of the two alleles of rs12923832, a functional study was initiated in cell cultures. The 676 bp fragment of *HERC2* intron 86 containing the DNA variation was subcloned together with a 514 bp fragment of human *OCA2* promoter elements into the vector pGL4.10. Gene expression array analyses indicated that

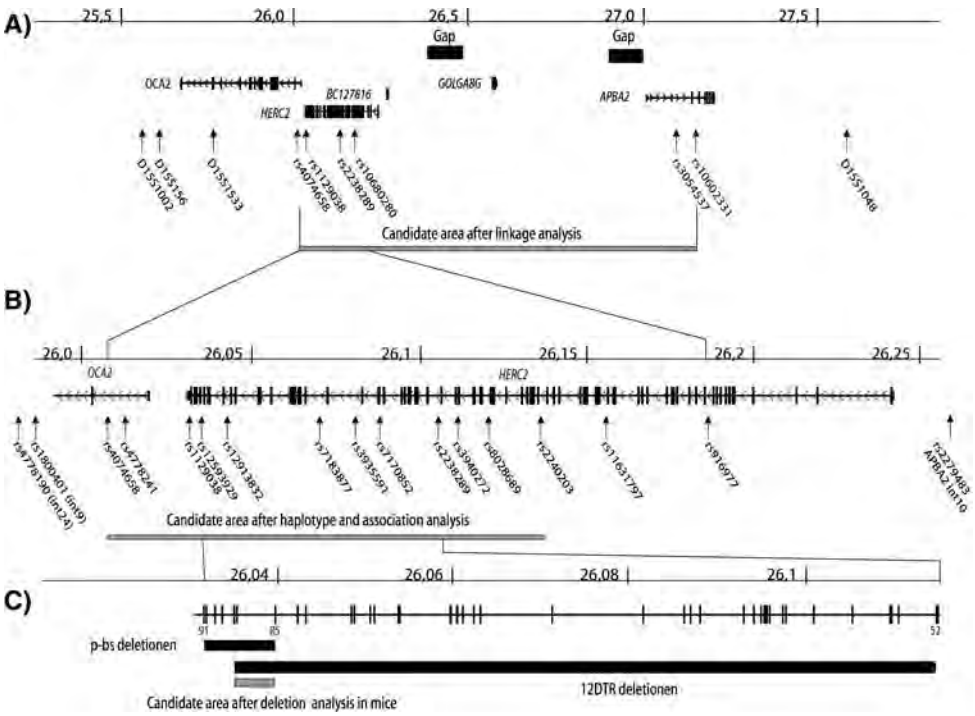

Fig. 3 The genetic and physical map of the human eye color locus. The map on chromosome 15q13.1 covers the physical map from positions 25.3 to 26.8 Mbp and includes DNA sequence gaps, genes and SNPs. The *gray* bars show the candidate areas found by the linkage studies (**a**), association study and the haplotype analysis (**b**) and deletions reported in mice (**c**)

Caco2 cells expresses *OCA2* mRNA in their differentiated state (http://gastro.imbg.ku.dk/chipchip/, searchterm = OCA2), we therefore transfected the human colon carcinoma cell line Caco2 and also the African Green Monkey SV40-transformed kidney fibroblast cell line COS7 with the *OCA2* promoter constructs (Fig. 4a). Both the cell lines were able to initiate reporter gene expression from the *OCA2* promoter. The *OCA2* promoter was approximately 10 times more active in Caco2 cells compared to COS7 cells. The 676 bp *HERC2* fragment including rs12913832 reduced the expression of the luciferase reporter gene by 60–70% in both Caco-2 and COS7 cells. In Caco2 cells, a significant difference in repressor activity between the A- (brown) and the G-allele (blue) was observed ($P < 0.05$ Fig. 4a). These results suggest a conserved regulatory element in the intron 86 sequence that acts as a transcriptional silencer with gene regulatory power in colon carcinoma cell which could be the same in melanocytes.

An electrophoretic mobility shift assay analysis of double stranded oligonucleotides carrying either the rs12913832*A or the *G-allele was performed using nuclear extract from differentiated Caco2 cells in order to investigate protein interactions to the rs12913832 region (Fig. 4b). Three specific protein/DNA complexes (Sc1, Sc2, Sc3 in Fig. 4b) were formed with the A-allele probe (lane 1), as these complexes can be competed by excess of unlabeled A-allele and G-allele oligonucleotides but not by an unspecific oligonucleotide (lane 2, 3, and 4). A similar result was obtained using the G-allele as probe (lane 5–6); however the binding pattern was different between the two variants indicating a differential binding of nuclear factors to the two variants. Especially the Sc3 complex is more pronounced using the G-allele probe compared to A-allele probe.

Eye color associations with hair color

Investigation of blond versus dark hair color in family CFG#694 demonstrated association of blond hair with the BEY1 phenotype and association of dark hair color with the BEY2 phenotype ($P = 0.0056$, Table 3). Analyzing the total material, the SNPs rs12593529*G and rs7495174*A were associated with the BEY2 phenotype (Table 3). A linkage study in this family using more than 400 markers versus dark or blond hair color resulted in an LOD score of $Z = 4.21$ at $\theta = 0.0$ to the marker D14S72 which is close to a possibly candidate gene *RABGGTA*, for hair color (GenBank acc. no. NM_182836).

Discussion

Blue/brown eye color is linked to a 166 Kbp region

Genes controlling the phenotypic variation of eye color from total brown versus blue eye color was previously mapped to chromosome 15q by a complete genome scan of five informative Danish families. The genome wide scan was expanded by a multipoint linkage analysis of additional 40 Danish families and resulted in an LOD score of $Z = 32.2$ (Eiberg and Mohr 1996) and excluded all other

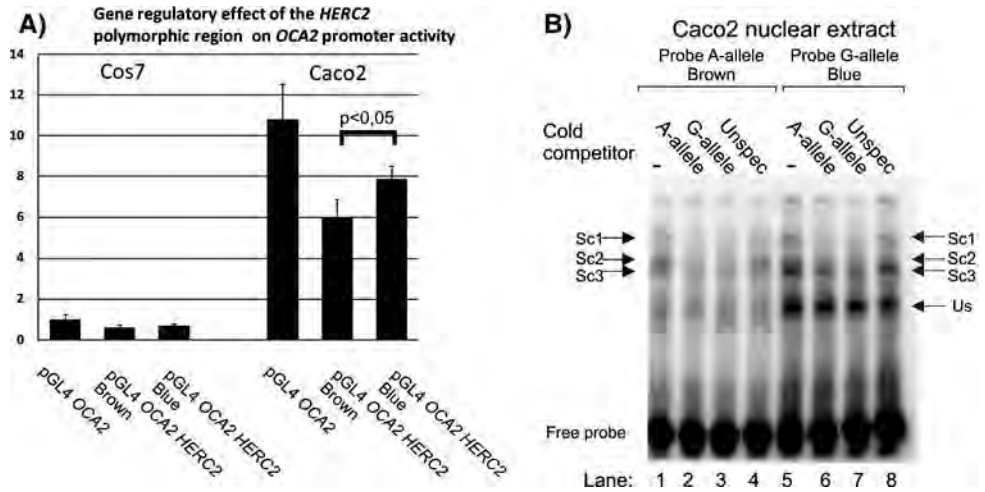

Fig. 4 Functional studies of the *OCA2* promoter regulation in COS7 and Caco2 cell cultures. **a** The *OCA2* promoter and *HERC2* intron sequences from blue and brown eyed individuals were inserted in a luciferase reporter gene plasmid (pGL4.10) and transfected into COS7 or Caco2 cell. The measured luciferase activities were corrected for transfection efficiency and normalized to the expression of pGL4-*OCA2* in COS7 cells, $N = 4$. *P*-values were calculated with Student *T*-test. **b** Electrophoretical mobility shift assay of the conserved element demonstrated binding of Caco2 cell nuclear factors to both the A-allele (brown eye) and G-allele (blue eye). Unlabeled oligonucleotides were added to demonstrate specific binding to both variants. Specific complexes are indicated by *Sc1-3*. An unspecific complex is indicated by *Us*

Table 3 Association between 5 SNPs and the two brown eye phenotypes BEY1 and BEY2

SNP	Genotype	Number of individuals Phenotype		Fisher exact test
		BEY1	BEY2	
rs1800401 OCA2 (ex9)	G,G	8 (17)	24 (73)	$P = 0.19$ ($P = 0.28$) Total $P = 0.06$
	G,A	2 (2)	1 (2)	
rs4778137 OCA2 (int1)	C,C	4 (5)	8 (28)	$P = 0.41$ ($P = 0.27$) Total $P = 0.33$
	C,A	6 (14)	17 (47)	
rs7495174 OCA2 (int1)	A,A	0 (0)	6 (16)	$P = 0.10$ ($P = 0.018$) Total $P = 0.0018$
	A,G	10 (19)	18 (58)	
rs12593529 HERC2 (int90) Haplotype-10	G,A	0 (0)	7 (17)	$P = 0.071$ ($P = 0.014$) Total $P = 0.0011$
	A,A	10 (19)	18 (58)	
rs3935591 HERC2 (int83)	A,G	6 (14)	14 (40)	$P = 0.72$ ($P = 0.07$) Total $P = 0.06$
	G,G	4 (5)	11 (35)	
Hair color in family CFB#694	Blond	10	0	$P = 0.00056$
	Brown	1	6	

The numbers of individuals for the two phenotypes BEY1 and BEY2 represent unrelated normal persons and the number in brackets represent the number of children up to the first 4 with brown eyes. Most significant associations for BEY1 and BEY2 are found to rs12593529 and hair color in family CFB#694

P-values were calculated using Fisher exact test

potential loci for the blue/brown eye color phenotype. The candidate locus was mapped to 15q12–13 flanked by the markers D15S156 and D15S144 with a Z_{max} close to marker D15S165. This locus included the *OCA2* gene, which was proposed as the candidate gene for the BEY2 phenotype.

In this study, we have minimized the candidate region for brown/blue iris pigmentation by linkage analysis to a 1.1 Mb interval flanked by the markers rs4074658 and rs10602331 and further to a 166 Kbp fragment flanked by the SNPs rs4074658 and rs2240203 by haplotype construction within the investigated trios. This fragment includes the *OCA2* intron 1 and the *OCA2* promoter and the intergenic region plus the 3′end of *HERC2* until intron 20 (Table 1; Fig. 3b).

A shared haplotype among blue eyed individuals is almost perfect and suggests the blue eye color phenotype is caused by a founder mutation

More than 97% of the analyzed persons with blue eyes carried the haplotype h-1 and the remaining 3% carried the haplotypes h-2, h-3 and h-4 (Table 2). The origin of these three haplotypes could be explained by recombination events or mutations younger in age than the original mutation. Seven unrelated individuals of Mediterranean origin with blue eyes carried the h-1 haplotype on both the chromosomes, suggesting this haplotype is identical for the blue eye individuals in other human populations.

rs12913832 is located within a regulatory element that inhibits expression of OCA2

Our association study suggests both rs1129038*A and rs12913832*G as preferable candidate mutations responsible for the blue eye color phenotype. The strong conservation across species propose the DNA region where rs12913832*G is located to have important regulatory functions. This is supported by the ESPERR regulatory potential presented by the UCSC genome browser (Fig. 5a). We showed that a 676 bp sequence, representing the *HERC2* intron 86 harboring rs12923832, reduced the expression of the luciferase reporter gene 60–70% both in Caco-2 cells and COS7 cells, when it was under control of the *OCA2* promoter (Fig. 4a). These results support a regulatory function of *OCA2* for this element and this regulatory function is conserved among species (Table 4). The rs1129038*A and rs12913832*G alleles demonstrated different regulatory effects on the *OCA2* promoter activity in Caco2 cells, and this is supported by the EMSA study where different binding affinities of the Caco2 nuclear factors to the blue and brown alleles were shown. This infers that the two alleles have different gene-regulatory activities in vivo for the two alleles; however, given the large differences in observed *OCA2* activity in two cell types used in the study, this would have to be addressed to experiments in human melanocytes.

Searching for the possible regulatory element in the conserved sequence revealed the transcriptional binding

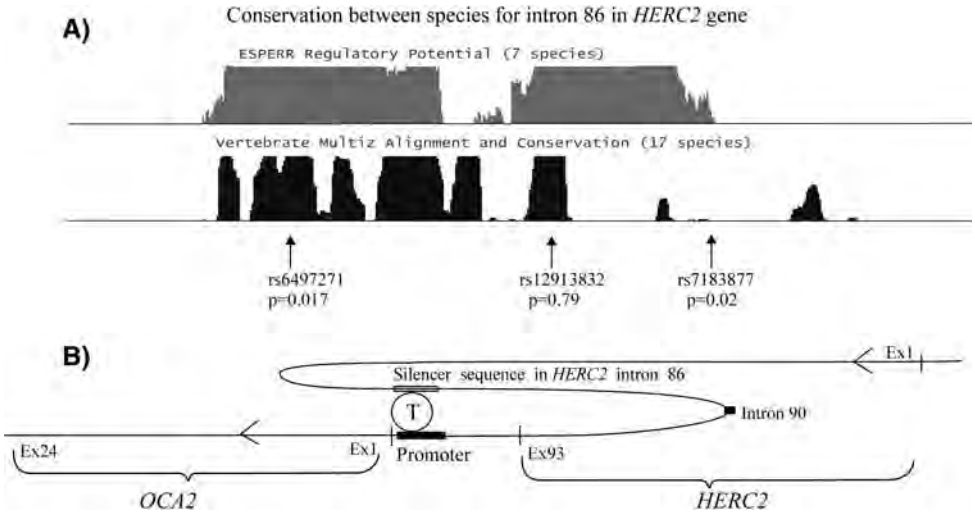

Fig. 5 Graphic presentation of the highly conserved region of *HERC2* intron 86. **a** The upper curve shows the ESPERR regulatory potentials and the lower part shows the area of conservations between species (adapted from UCSC Human Genome Browser). Three different SNPs are represented in the region but only the rs12913832*G allele is 100% associated with the blue allele ($P = 0.792$ in the CEPH material from the UCSC genome browser and HapMap CEU). **b** A hypothetic model for the silencer-promoter complex (*T*) for regulation of the *OCA2* gene activity. The rs12593929*G allele in intron 90 is associated with the BEY2 phenotype (Table 4) and may interfere with the binding complex activity

Table 4 Conservation of the eye color silencer sequence in the *HERC2* intron 86 in 9 different species

Species	Eye color	DNA-Library	DNA sequence
Homo	Blue	hg18_dna	TTCATTTGAGCATTAA**G**TGTCAAGTTCTGCACGCTAT
Homo	Brown	hg18_dna	TTCATTTGAGCATTAA**A**TGTCAAGTTCTGCACGCTAT
Chimpanse	Brown	panTro2_dna	TTCATTTGAGCATTAA**A**TGTCAAGTTCTGCACGCTAT
Rhesus monkey	Brown	rheMac2_dna	TTCATTTGAGCATTAA**A**TGTCAAGTTCTGCACGCTAT
Horse	Brown	equCab1_dna	TTCACTTGCACGCTAA**A**TGTCAAGTGCTGCACAATGT
Cow	Brown	bosTau2_dna	TTCATCTGCATGGTAA**A**TGTCAAGTAC-ACACACTGT
Cat	Brown-yellow	felCat3_dna	TTCATTTGCATGTTAA**A**TGTCAAGTACCACACAATAC
Dog	Brown-yellow	canFam2_dna	TTCATTTGCATGTTAA**A**TGTCAAGTGC-ACACAATAT
Rat	Brown	rn4_dna	TTCATTTGCGTATTAA**A**TGTCAA
Mouse	Brown	Mm8_dna	TTCATTTGCGTATTAA**A**TGTCAAATGCCATGCACTAT
Consensus sequence - blue eye			Ttca-ttg-----taa**G**tgtcaa-t-c----c-tat
Consensus sequence - brown eye			Ttca-ttg-----taa**A**tgtcaa-t-c----c-tat
Nkx-2.5 target site; match allele for blue eye color			TYAA**G**TG
CdxX-1 target site; match allele for brown eye color			YAKWAWW

The DNA sequences for the 9 species are from the UCSC genome browser (May 2006). The grey shaded sequences represent the binding site region and the bold G or A nucleotide in the gray area represent the variation found between blue and brown eye color. Nkx-2.5 match the blue sequence (score 0.99) and CdxX-1 match brown (score 0.92) and blue sequences (score 0.87) analyzed by TFSEARCH

sequence TAA**A**TGTCAA (match 0.92, the rs12923832 A-allele is in bold) for the homeodomain transcription factor Cdx-1, and the corresponding G-allele results in a swift binding to the homeodomain transcription factor Nkx-2.5 (TAA**G**TGTCAA, match 0.98, the G-allele is in bold). Since Cdx-1 and Nkx-2.5 are expressed in the intestine and the heart, respectively, it is not likely that they are involved in the iris melanocytes development and regulation. This

indicates that iris melanocytes express other homeodomain transcription factors involved in the regulatory activity of the conserved rs12923832 region. These differences in the binding affinity to nuclear factor by a G to A change in the recognition sequence can explain why the silencer-promoter complex involved in regulation of *OCA2* gene can change conformation and thereby activity (Heinemeyer et al. 1998; Carter et al. 2002).

The SNP rs12593529*G allele associated with the BEY2 phenotype (Table 3) is located in *HERC2* intron 90 between the *OCA2* promoter and rs12923832 (Fig. 5b). This SNP may have a stabilizing or destabilizing effect on the proposed *OCA2* silencer-promoter complex. Interestingly, the rs7495174*A in *OCA2* intron 1 was associated to rs12593529*G (Table 3) and suggest a common haplotype for approximately 30 % of individuals with BEY2 phenotype.

Different brown eye color phenotypes

The SNP rs12913832 is found to be associated with the brown and blue eye color, but this single DNA variation cannot explain all the brown eye color variation from dark brown over hazel to blue eyes with brown spots.

All individuals carrying the BEY1 and BEY2 phenotypes in family CFG#694 (Fig. 2) shared the same two haplotypes h-5 and h-1. Further, all the brown-eyed individuals shared a 7.6 Mbp region that includes the entire *OCA2* coding region from D15S113 to D15S144 (chr15: 23.8–31.4 Mbp) but excludes any regulatory sequences located in *OCA2* exon 1 and upstream promoter region for the brown eye and brown hair color variations in this family. A recent study by Duffy et al. (2007) showed 90% association with the three SNPs alleles (TGT) in *OCA2* intron 1 for human blue eye color. This region may have a regulatory function in the brown eye color variation, but is more likely that the TGT haplotype for these SNPs is associated with the h-1 haplotype identified for the *HERC2* gene found in this study.

However, our data for family CFB#694 were not in accordance with brown variations controlled by alleles in *OCA2* intron 1, since persons having different brown eye phenotypes carry identical *OCA2-HERC2* haplotypes.

Studies of the BEY1 and BEY2 phenotypes showed association between blond hair and BEY1 and association between dark hair color and BEY2 ($P = 0.0056$, Table 3). *RABGGTA* on chromosome 14 is the closest candidate gene to D14S72 (an LOD score $z = 4.21$ for $\theta = 0.0$) and this gene may moderate the brown eye phenotype in family CFG#694 as shown in Table 3. Mutation in *RABGGTA* gene is known to result in albinism in mice (Novak et al. 1995).

Association or linkage between BEY2 and /HCL3 (brown hair color) has been described (Eiberg and Mohr 1996; Duffy et al. 2007), but the results point toward other loci unlinked to *OCA2-HERC2* (Duffy et al. 2007). Candidates for these loci are the *MATP* and *MC1R* where association to human hair color has been found (Remenyi et al. 2004). Interestingly also the BEY2 phenotypes shows association with rs7495174*A ($P = 0.0018$) and rs12593529*G ($P = 0.0010$) in the *OCA2-HERC2* loci (Table 3) which strongly suggests that DNA variation in this region is associated with the BEY2 phenotype (Eiberg and Mohr 1996).

The origin of the founder mutation

The mutations responsible for the blue eye color most likely originate from the neareast area or northwest part of the Black Sea region, where the great agriculture migration to the northern part of Europe took place in the Neolithic periods about 6–10,000 years ago (Cavalli-Sforza et al. 1994).

The high frequency of blue-eyed individuals in the Scandinavia and Baltic areas indicates a positive selection for this phenotype (Cavalli-Sforza et al. 1994; Myant et al. 1997). Several theories has been suggested to explain the evolutionary selection for pigmentation traits which include UV expositor causing skin cancer, vitamin D deficiency, and also sexual selection has been mentioned. Natural selection as suggested here makes it difficult to calculate the age of the mutation.

Conclusion

In conclusion, we have identified a conserved regulatory element within intron 86 of the *HERC2* gene that is perfectly associated with the brown/blue eye color in studied individuals from Denmark, Turkey and Jordan. This element had an inhibitory effect on the *OCA2* promoter activity in cell cultures, and the blue and the brown alleles were shown to bind non-identical subsets of nuclear extracts. In total, all these data strongly support a model where the blue eye color in humans is caused by homozygosity of the rs12913832*G allele.

Web Resources

The accession numbers and URLs for data presented herein are as follows:

AceView NCBI, http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html (for functional information of *OCA2* and *HERC2*)

ClustalW, http://www.ebi.ac.uk/clustalw/ (for alignment of sequences)

Fishers exact test, http://home.clara.net/sisa/fisher.htm (test for significant genotype-phenotype distributions)

GenBank dbSNP, http://genome.ucsc.edu/cgi-bin/hgGateway (for *Homo sapiens* chromosome 15 genomic contig [assession number NT_010280]) *OCA2* (NM_000275), *HERC2* (NM_004667)

HapMap SNPs, http://genome.cse.ucsc.edu/cgi-bin/hgTracks (for SNPs in *OCA2* and *HERC2* genes)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omin/ (for *OCA2* and *HERC2*).

Primer3, http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi (for primer construction)

TFSEARCH, http://www.cbrc.jp/research/db/TFSEARCH.html (for determination of binding sites for a transcription factor to a specific DNA sequence)

The Intestinal Transcription Factor Target Database http://gastro.imbg.ku.dk/chipchip/, searchterm = OCA2

UCSC Human Genome Browser (July 2003 and May 2006) http://genome.cse.ucsc.edu/cgi-bin/hgGateway?org=human (for detections of SNPs, conserved regions, exon-introns in *OCA2*, *HERC2* gene and chromosome 15).

Acknowledgments We would like to thank Karin U. Hansen for technical assistance. The project was financially supported by the Danish Research Foundation (JN 22-02-0208). Wilhelm Johannsen Centre is established by the National Danish Research Center.

References

Ancans J, Tobin DJ, Hoogduijn MJ, Smit NP, Wakamatsu K, Thody AJ (2001). Melanosomal pH controls rate of melanogenesis, eumelanin/ phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. Exp Cell Res 268:26–35

Carter D, Chakalova L, Osborne CS, Dai Y-F, Fraser P (2002). Long-range chromatin regulatory interactions in vivo. Nat Genet 32:623–626

Cavalli-Sforza LL, Menozzi P, Piazza A (1994). The History and geography of Human genes. Princeton University Press, Princeton

Duffy DL, Box NF, Chen W, Palmer JS, Montgomery GW, James MR, Hayward NK, Nicholas G, Martin NG, Sturm RA (2007). A three-single-nucleotide polymorphism haplotype in intron 1 of *OCA2* explains most human eye-color variation. Am J Hum Genet 80:241–252

Eiberg H, Mohr J (1996). Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. Eur J Hum Genet 4:237–241

Eiberg H, Nielsen LS, Klausen J, Dahlén M, Kristensen M, Bisgaard ML, Møller N, Mohr J (1989). Linkage between serum cholinesterase 2 (CHE2) and δ-crystalline gene cluster (CRYG): assignment to chromosome 2. Clin Genet 35:313-321

Frudakis T, Thomas M, Gaskin Z, Venkateswarlu K, Chandra KS, Ginjupalli S, Gunturi S, Natrajan S, Ponnuswamy VK, Ponnuswamy KN (2003). Sequences associated with Human iris pigmentation. Genetics 165:2071–2083

Frudakis T, Terravainen T, Thomas M (2007) Multilocus *OCA2* genotypes specify human iris colors. Hum Genet 122:311–326. doi:10.1007/s00439-007-0401-8

Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatieva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA et al (1998) Databases on transcriptional regulation: TRANSFAC, TRRD, and COMPEL. Nucleic Acids Res 26:364–370

Lamoreux ML, Zhou BK, Rosemblat S, Orlow SJ (1995) The pin-keyed-dilution protein and the eumelanin/pheomelanin switch: in support of a unifying hypothesis. Pigment Cell Res 8:263–270

Lee ST, Nicholls RD, Jong MT, Fukai K, Spritz RA (1995). Organization and sequence of the human *P* gene and identification of a new family of transport proteins. Genomics 20:354–63

Lehman AL, Nakatsu Y, Ching A, Bronson RT, Oakey RJ, Keiper-Hrynko N, Finger JN, Durham-Pierre D, Horton DB (1998) A very large protein with diverse functional motifs is deficient in rjs (runty, jerky, sterile) mice. Proc Natl Acad Sci USA 95:9436–9441

Myant NB, Fobes SA, Day INM, Gallagher J (1997) Estimation of the age of Ancestral Argenine$_{3500}$ → Glutamine mutation in Human ApoB-100. Genomic 45:78–87

Novak EK, Reddington M, Zhen L, Stenberg PE, Jackson CW, McGarry MP, Swank RT (1995) Inherited thrombocytopenia caused by reduced platelet production in mice with the gunmetal pigment gene mutation. Blood 85:1781–1789

Posthuma D, Visscher PM, Willemsen G, Zhu G, Martin NG, Slagboom PE, de Geus EJ, Boomsma DI (2006). Replicated linkage for eye color on 15q using comparative ratings of sibling pairs. Behav Genet 36:12–17

Puri N, Gardner JM, Brilliant MH (2000). Aberrant pH of melanosomes in pink-eyed dilution (*P*) mutant melanocytes. J Invest Dermatol 115:607–613

Rebbeck TR, Kanetsky PA, Walker AH, Holmes R, Halpern AC, Schuchter LM, Elder DE, Guerry D (2002). P gene as an inherited biomarker of human eye color. Cancer Epidemiol Biomarkers Prev 11:782–784

Remenyi A, Scholer HR, Wilmanns M (2004). Combinatorial control of gene expression. Nat Struct Mol Biol 11:812–815

Russell LB, Montgomery CS, Casheiro NLA, Johnaso DK (1995) Complementation analysis of 45 mutations encompassing the pink-eyed dilution locus of the mouse. Genetics 141:1547–1562

Schäffer AA, Gupta SK, Shriram K, Cottingham RW Jr (1994). Recomputation in linkage analysis. Hum Hered 44:225–237

Spritz RA, Bailin T, Nicholls RD, Lee ST, Park SK, Mascari MJ, Butler MG (1997). Hypopigmentation in the Prader-Willi syndrome correlates with P gene deletion but not with haplotype of the hemizygous P allele. Am J Med Genet 71:57–62

Sturm RA, Teasdale RD, Box NF (2001) Human pigmentation genes; identification, structure and consequences of polymorphic variation. Gene 277:49–62

Toyofuku K, Valencia JC, Kushimoto T, Costin GE, Virador VM, Vieira WD, Ferrans VJ, Hearing VJ (2002). The etiology of oculocutaneous albinism (OCA) type II: the pink protein modulates the processing and transport of tyrosinase. Pigment Cell Res 5:217–224

Troelsen JT, Mitchelmore C, Olsen J (2003a) An enhancer activates the pig lactase phlorizin hydrolase promoter in intestinal cells. Gene 305:101–111

Troelsen JT, Olsen J, Møller J, Sjöström H (2003b). An upstream polymorphism associated with lactase persistence has increased enhancer activity. Gastroenterology 125:1686–1694

Yonggang J, Rebert NA, Joslin JM, Higgins J, Schultz RA, Nicholls RD (2000). Structure of the highly conserved *HERC2* gene and of multiple partially duplicated paralogs in Human. Genome Res 10:319–329

Zhu G, Evans DM, Duffy DL, Montgomery GW, Medland SE, Gillespie NA, Ewen KR, Jewell M, Liew YW, Hayward NK, et al (2004). A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q. Twin Res 7:197–210