HHS Public Access

Author manuscript

Hum Mutat. Author manuscript; available in PMC 2022 October 01.

Published in final edited form as:

Hum Mutat. 2021 October ; 42(10): 1239–1253. doi:10.1002/humu.24257.

A custom capture sequence approach for oculocutaneous albinism identifies structural variant alleles at the *OCA2* locus

Stacie K. Loftus[1], **Linnea Lundh**[2], **Dawn E. Watkins-Chow**[1], **Laura L. Baxter**[1], **Erola Pairo-Castineira**[3], **NISC Comparative Sequencing Program**[4], **Ian J. Jackson**[3], **William S. Oetting**[5], **William J. Pavan**[#1], **David R. Adams**[#6]

[1]Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

[2]Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 10 Center Drive, MSC 1851, Bethesda, MD, 20892, USA.

[3]Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK.

[4]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

[5]Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, Minnesota 55455, USA.

[6]Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, 10 Center Drive, MSC 1851, Bethesda, MD, 20892, USA.

[#] These authors contributed equally to this work.

Abstract

Oculocutaneous albinism (OCA) is a heritable disorder of pigment production that manifests as hypopigmentation and altered eye development. Exon sequencing of known OCA genes is unsuccessful in producing a complete molecular diagnosis for a significant number of affected individuals. We sequenced the DNA of individuals with OCA using short-read custom capture sequencing that targeted coding, intronic and non-coding regulatory regions of known OCA genes and GWAS-associated pigmentation loci. We identified an *OCA2* complex structural variant (CxSV), defined by a 143kb inverted segment reintroduced in intron 1, upstream of the native location. The corresponding CxSV junctions were observed in 11/390 probands screened. The 143kb CxSV presents in one family as a copy number variant (CNV) duplication for the 143kb

Corresponding author: Stacie Loftus, Ph.D., Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD USA 20892, Tel: 571-235-8853, sloftus@mail.nih.gov.

Web resources

Population frequencies For SNVs were obtained though dbSNP (https://www.ncbi.nlm.nih.gov/snp/). Linkage disequilibrium values and haplotype frequencies for 1000 Genomes were obtained using the LDlink suite of tools (https://ldlink.nci.nih.gov/). Evaluation of *OCA2* intron 1 features from previous studies was performed by ReMap2020 at (http://remap.univ-amu.fr/genome_tracks_page). Variant details and approved nomenclature have been deposited in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/variation/1098719/ and https://www.ncbi.nlm.nih.gov/clinvar/variation/1098720/).

region. In the remaining 10/11 families, the 143kb CxSV acquired an additional 184kb deletion across the same region, restoring exons 3–19 of *OCA2* to a copy-number neutral state. Allele-associated haplotype analysis found rare SNVs rs374519281 and rs139696407 are linked with the 143kb CxSV in both *OCA2* alleles. For individuals in which customary molecular evaluation does not reveal a biallelic OCA diagnosis, we recommend preliminary screening for these haplotype-associated rare variants, followed by junction-specific validation for the *OCA2* 143kb CxSV.

Keywords

albinism; *OCA2*; inversion; melanosome; pigmentation; copy number neutral

Introduction

Oculocutaneous albinism (OCA) is an autosomal recessive disorder of considerable phenotypic and genetic heterogeneity, with a wide range of severity for the associated clinical manifestations. OCA-related hypopigmentation affects the skin, hair and eyes (King et al. 2003; Marçon and Maia 2019), and skin hypopigmentation puts affected individuals at increased risk of UV-induced skin cancers, especially in settings where sun protection is not a customary practice (Hawkes et al. 2013; Awe and Azeke 2018; Nathan et al. 2019; Rayner et al. 2020). OCA-associated changes in eye development result in reduced visual acuity, principally due to foveal hypoplasia. Additional consequences include nystagmus and optic nerve mis-routing (Creel et al. 1990; Oetting et al. 1996; Kruijt et al. 2018). Alterations of 8 genes/loci have been associated with non-syndromic albinism: *TYR* (OCA types 1A and 1B), *OCA2* (OCA type 2), *TYRP1* (OCA3), *SLC45A2* (OCA4), *SLC24A5* (OCA6), *C10orf11*/*LRMDA* (OCA7), *DCT* (OCA8) (Montoliu et al. 2014; Pennamen et al. 2020) and a linkage region on 4q24 identified in a single consanguineous family (OCA5) (Kausar et al. 2013). Three of these genes encode melanosomal enzymes involved in the production of melanin pigment (*TYR, TYRP1, DCT*) and three encode transmembrane proteins that modulate melanosomal pH (*OCA2, SLC45A2, SLC24A5*). OCA may also be a component of multi-system syndromes, including Hermansky-Pudlak Syndrome (HPS), Chediak-Higashi Syndrome (CHS), or Griscelli Syndrome (Pastural et al. 1997; Anikster et al. 2002; Ménasché et al. 2003; Montoliu et al. 2014). Along with the 8 OCA loci, 14 additional genes have been discovered for syndromic OCA.

The customary estimate of albinism prevalence worldwide is ~1/17,000 (King and Summers 1988; Montoliu et al. 2014), with allelic variants primarily residing in either the *TYR* or *OCA2* loci (Lasseaux et al. 2018). However, the frequency of OCA and OCA-associated alleles varies among geographically distinct populations. For example, *TYR* alleles have been found responsible for between 40–70% of OCA individuals of European descent (Hutton and Spritz 2008; Lasseaux et al. 2018). In contrast, the estimated rate of OCA is between 1/1,000 to 1/15,000 in populations from Sub-Saharan Africa (Kromberg et al. 1989; Venter et al. 1995), Tanzania (Luande et al. 1985), Nigeria (Okoro 1975), Zimbabwe (Lund 1996), and Cameroon (Aquaron 1990). This has been attributed in part to a founder effect from an *OCA2* allele (which harbors a 2.7 kilo base pair (kb) deletion of exon 7 and flanking intronic sequences) segregating within these populations at a carrier rate as high as 1 in

16 (Durham-Pierre et al. 1994; Stevens et al. 1995; Puri et al. 1997; Stevens et al. 1997; Mitchell and Reigada 2008).

OCA2 encodes P protein (hereafter named OCA2), a 12-transmembrane domain protein that shares homology with anion transporters and is proposed to function as a melanosomal membrane channel (Rosemblat et al. 1994; Lee et al. 1994b; Sitaram et al. 2009; Bellono et al. 2014). *OCA2* was originally identified as the human homolog of mouse pink-eyed dilution (*p*), one of the first genes associated with hypopigmentation (Lyon et al. 1992; Rinchik et al. 1993). Decreased OCA2 expression levels reduce the number of mature, stage 4, pigmented melanosomes, while increasing the number of stage 2 melanosomes (Park et al. 2015). While the biochemical function for OCA2 is not completely understood, it has been suggested that OCA2 regulates pH in stage 1 and stage II melanosomes by modulating chloride currents (Rosemblat et al. 1998; Orlow and Brilliant 1999; Bellono et al. 2014; Le et al. 2020). Taken together, these studies highlight the role of OCA2 in maintaining melanosome pH, which is critical for tyrosinase enzyme activity and melanin production (Ancans et al. 2001; Le et al. 2020).

Genome Wide Association Studies (GWAS) identify Single Nucleotide Variants (SNVs) segregating in populations that are associated with clinical phenotypes. Multiple GWAS have identified the *OCA2* locus to be a major contributor to pigmentation variation in human hair, skin, and eye color in individuals of European, African and Latin American descent. In total, 15 distinct GWAS studies have documented 43 allelic-trait associations at the *OCA2*/*HERC2* locus, including pigmentation (Han et al. 2008; Zhang et al. 2013; Crawford et al. 2017; Morgan et al. 2018; Visconti et al. 2018; Adhikari et al. 2019; Rashkin et al. 2020), sunburn (Liu et al. 2015), nevus count (Landi et al. 2020), melanoma (Law et al. 2015; Ransohoff et al. 2017; Landi et al. 2020), and non-melanoma skin cancer (Asgari et al. 2016; Chahal et al. 2016; Liyanage et al. 2019; Sarin et al. 2020). Therefore, common SNVs at *OCA2* have a demonstrable impact on OCA2 protein function and expression levels, affecting not only individuals with albinism-related *OCA2* alleles but also broadly impacting pigment variation and skin cancer disease risk across populations worldwide.

A variety of exon-targeted sequencing approaches have been critical in identifying OCA variant alleles in diverse populations. However, 10–25% of OCA individuals remain without a definitive, bi-allelic diagnosis (Wei et al. 2010; Simeonov et al. 2013; Kruijt et al. 2018; Lasseaux et al. 2018; Grønskov et al. 2019; Zhong et al. 2019; Okamura and Suzuki 2020). Included in this subset are many individuals for whom only one deleterious allele is identified. Of note, current exon-based screening could miss alterations in non-coding regulatory regions or gross locus rearrangements that do not alter exon number. *OCA2* appears particularly prone to rearrangements, as the OCA literature contains reports of many large-scale structural variants at the *OCA2* locus, with junctions mapping to highly repetitive Alu and L1 genomic repeats within *OCA2* introns (Shahzad et al. 2017; Lasseaux et al. 2018). In this study, we sought to identify additional alleles present in OCA individuals without a definitive bi-allelic diagnosis by utilizing custom capture sequencing (CCS) to target the genomic regions of a select number of albinism and pigmentation loci. We identified 11 unrelated families that harbored identical junction sequences defining a complex structural variant (CxSV) at *OCA2* by applying a combined approach of CCS

in 211 OCA probands, followed by CxSV junction-target screening in 179 additional probands. Detailed analysis of CxSVs in these families provided evidence of an initial founder allele that subsequently underwent two structural alterations, giving rise to two CxSVs that are now segregating among the population. These studies highlight the need for more comprehensive genomic analyses of OCA loci during molecular diagnostic workup, either routinely or using a staged approach.

Materials and Methods

Editorial Policies and Ethical Considerations

Individuals were drawn from a prior NIH Clinical Center natural history study (IRB approved study 2009-HG-0035), as well as a de-identified cohort donated to the NIH by Dr. Richard King and Dr. William Oetting at the University of Minnesota consisting of DNA samples from individuals diagnosed with OCA under IRB-9508M10178. This study conforms to recognized standards in the US Federal Policy for the Protection of Human Subjects.

OCA Patients

The individuals included in the sequencing cohort were selected based on the lack of a definitive molecular diagnosis, despite some degree of exon-targeted sequencing of OCA loci. Prior Sanger sequencing of exons varied across the population, in some cases only including the most common loci (*TYR* and *OCA2*). An initial cohort of 221 individuals was screened using custom capture sequencing, of which 211 individuals represented probands and the remaining 10 represented related family members with OCA. Following the initial detection of CxSV junctions in 7 families by CCS, a separate cohort of 179 OCA probands was screening using a combination of TaqMan assays for rare variants rs374519281 and rs139696407 followed by PCR amplification of novel CxSV junction fragments. This process identified an additional four families. Thus a total of 400 separate OCA individuals, corresponding to 390 probands, were screened for the identified *OCA2* CxSVs. Clinical presentations for the OCA individuals identified with *OCA2* CxSVs is in Supp. Table S3.

For individuals seen at the NIH Clinical Center, DNA was isolated from whole blood drawn into BD Vacutainer® EDTA tubes. Genomic DNA (gDNA) was isolated using the Gentra Puregene Blood Kit protocol (Qiagen, CN: 158389). Quantification of gDNA was via NanoDrop™ ND1000 Spectrophotometer and/or Quant-IT™ PicoGreen™ dsDNA assay kits (Invitrogen, CN: P7589) with SpectraMax M5 microplate reader and SoftMax(R) Pro 7 software.

Custom Capture Sequencing

211 OCA probands and a selected set of 10 additional affected family members were sequenced through short read CCS. The CCS library of bait probes was selected by first identifying 7,240,072 base pairs (bp) of unique sequence corresponding to 37 genes for OCA, syndromic albinism and GWAS-identified pigmentation phenotypes. This included 2.2 Mb exonic, intronic and non-coding regulatory regions at 6 OCA genes (*TYR*, *OCA2*, *TYRP1*, *LRMDA*, *SLC24A5*, and *SLC45A2*) and an additional 4.8 Mb of exonic and

non-coding regulatory features for 31 additional genes previously linked to syndromic albinism and/or human hair, skin and eye pigment variation (Morgan et al. 2018). From these coordinates we selected 58,080 unique bait probes 120 bp in length, which provide 93.5% coverage of regions identified, and span a total of 6,696,663 bp (Twist Bioscience) (Supp. Table S2).

Custom capture, paired-end, indexed libraries and short-read Illumina sequencing were generated at the NIH Intramural Sequencing Center (NISC). Sequencing libraries with ~325 base inserts were prepared for each sample from 100 ng of gDNA using the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences). Libraries were barcoded with unique dual indexes to minimize barcode hopping and pooled in groups of 16 in an equimolar ratio for capture enrichment. CCS was performed using the custom-designed bait probes from Twist Bioscience and the IDT xGen Universal Blockers-TS Mix according to the “EF Workflow” from Twist Bioscience. Typically, six captured pools were combined and sequenced on a NovaSeq 6000 (Illumina, Inc.) using version 1 chemistry. At least 7 million, paired-end 151 bp reads were obtained for each sample. Data was processed using RTA ver. 3.4.4.

Alignment and Genotype Calling

Reads were mapped to NCBI build GRCh37/hg19 using NovoAlign V3.02.07. The aligned lane BAM files were merged, sorted, and indexed. Duplicate sequence reads were removed with Samtools. The alignments were stored in BAM format, and then provided as input to bam2mpg (http://research.nhgri.nih.gov/software/bam2mpg/index.shtml), which calls genotypes at all covered positions using a probabilistic Bayesian algorithm (Most Probable Genotype, or MPG). The genotype calls were compared against Illumina Human 1M-Quad genotype chips, and genotypes with an MPG score of 10 or greater showed >99.89% concordance with SNP Chip data. Sequence bases with Phred quality score of less than 20 (Q20) were ignored. Only reads with mapping quality greater than 30 were included for the analysis. The normal orientation for paired-end reads relative to the reference genome is left/right (LR). Paired-end read orientations that are annotated left/left (LL) and right/right (RR) denote read pairs that are indicative of an inversion relative to reference sequence.

SNV Variant Annotation

Variants were annotated using Annovar (http://www.openbioinformatics.org/annovar/annovar_gene.html). A number of filtering and prioritization steps were applied to reduce the number and to identify potentially pathogenic variants, similar to the methods used in previous studies (Ng et al. 2009; Ng et al. 2010). Missense variants were sorted by the degree of severity of functional disruption prediction using CDPred. Variants detected in dbSNP (version 137), 1000 Genomes, NHLBI 6500ESP, Polyphen, EXAC, CAD and HGMD were annotated.

Complex Structural Variant Assessment

BAM files were evaluated using Integrative Genomics Viewer (Broad Institute) (Robinson et al. 2017) to identify potential structural rearrangements. FASTA files were generated for paired-end reads that directly flanked altered orientation (LL and RR) or predicted

deletion junctions and were separately reassembled using Sequencher 5.4.6 (GeneCodes, Ann Arbor, MI) to identify the full sequence of the junction fragments. Bigwig files were generated using deepTools2 (Ramírez et al. 2016) and viewed using the UCSC browser, (http:genome.ucsc.edu/). All figure images of *OCA2* are presented relative to *OCA2* 5'to 3' orientation, rather than that of the reference genome orientation.

Predicted CxSV junctions and pathogenic variant predictions were verified by PCR amplification followed by Sanger sequencing in OCA individuals and corresponding family members. Following PCR amplification, bands were isolated and purified using QIAquick Gel Extraction Kit (Qiagen, CN: 28706) or ExoSAP-IT™ Express PCR Product Cleanup Reagent (Applied Biosystems, CN: 75001), and sequenced using BigDye™ Terminator Cycle Sequencing Kit (Applied Biosystems, CN: 4337458) or by Macrogen USA. PCR products were sent to Psomagen Service Center (Rockville, MD) for sequencing. PCR primer sequences used to amplify the novel Junctions 1, 2, and 3 are in Supp. Table S1, and amplified junction sequences including novel inserted sequences identified in family probands and in family members are in Supp. Figure S1. Junction 1 presents with a single, novel "g" insertion. Validation for Junction 2/LL was performed by two primer sets (JCT2A and JCT2B) as the sequences surrounding Junction 2 encoded non-unique sequences flanking a novel 31bp insertion. A primer including a portion of the novel 31bp insertion sequence was used as an anchor primer in each direction as it was unique to the complex rearrangement. The JCT2A primer set sequenced towards the *OCA2* 5' end, while the JCT2B primer set sequenced towards the *OCA2* 3' end.

Allele-specific Genotyping and Phase Transmission of SNV Variants

Genotyping of SNVs in family trios was performed using TaqMan SNP genotyping assays (Themofisher) (Supp. Table S5). All reactions were performed using TaqMan Universal PCR master mix under fast cycle conditions of StepOne™ Real-Time PCR system (Thermofisher). TaqMan assays for rare variants associated with the CxSV alleles, rs139696407 and rs374519281, were used to prescreen the second OCA patient population (179 individuals) prior to phase assessment and evaluation of junction fragment confirmation. Identification of a second, *trans* deleterious *OCA2* allele for Families 7 and 8 was done by PCR amplification followed by Sanger sequencing of *OCA2* exons 18 and 9, respectively. Primers used are supplied in Supp. Table S1.

Haplotype Assessment

CCS identified a total of 30/221 individuals that possessed the p.Val443Ile allele. Using the genotypes for all 30 OCA individuals, a single consensus haplotype, defined by the presence of least one shared allele among all 30 individuals, was obtained for a 46,852 bp region from intron 6 to intron 14 (chr15:28218003–28264855, Supp. Table S6) of the *OCA2* locus. The p.Val443Ile allele was present in *trans* for 4 individuals with the 143kb;184kb CxSV allele. Therefore, the p.Val443Ile haplotype was used to anchor the phase for 143kb;184kb CxSV allele in individuals carrying the p.Val443Ile allele in *trans*.

For the phase assessment of the two 143 kb segments (inverted and native) located on the 143kb-F11 CxSV allele, both the p.Val443Ile haplotype and the allele frequency within the

duplicated region for the proband in Family 11 were used in combination. For example, if a given p.Val443Ile-associated haplotype SNV was found at ~34% frequency, then the alternate SNV allele present at ~66% frequency would be identical on both regions of the 143 kb duplication. In contrast, if the p.Val443Ile haplotype SNV was found at ~66% frequency, this necessitated that 1) the one other copy of this ~66% SNV allele would be on the 143kb-F11 CxSV allele once, and 2) for that given SNV, the 143kb-F11 CxSV allele contained two duplicated regions each with a different haplotype. Linkage disequilibrium and population specific haplotype frequencies were calculated using LD link (https://ldlink.nci.nih.gov/) (Machiela and Chanock 2015).

Results

Targeted genomic sequence analysis reveals complex rearrangements in *OCA2*

In OCA individuals from seven families, CCS and IGV analysis identified 2 regions (Junction 1 and Junction 2) at the *OCA2* locus in which paired-end reads mapped with an insert size distinct from, and in a reverse orientation to that of the reference genome. The orientation and alignment of read pairs suggested an *OCA2* complex variant (CxSV) consistent with a 143 kb inverted segment of *OCA2* having been reinserted into intron 1, roughly ~33 kb away from the native location (Figure 1). This CxSV is defined by both Junction 1/RR, which spans sequence in introns 1 and 19 (blue arrows in Figure 1 A, C), and Junction 2/LL, which spans sequence in introns 1 and 2 (teal arrows in Figure 1 A, B) in addition to a deletion of 1680 bp (~1.7 kb) of intron 1 (Figure 1A, D). Further analyses of the CCS data identified two distinct altered read coverage profiles (Figure 1D) associated with these junction fragments, one present in 6 probands and 1 OCA-affected sibling from Families 1–6, and a second distinct profile in Family 11 (described below). Taken together, the position and sequence of the Junction 1 and 2 paired-end read alignments are consistent with a 143 kb fragment spanning intron 2 through intron 19 having been inverted and reinserted into intron 1, concomitantly creating a 1.7 kb deletion (referred to now as “143kb CxSV”) (Figure 2).

In addition to the 143 kb inverted segment, all OCA-affected probands in Families 1–6 showed evidence of a second structural variant, encompassing a 184 kb deletion, with boundaries defined by Junction 3 paired-end reads which mapped to unexpectedly distant locations in intron 2 and intron 19 (red arrows in Figure 1B, C). A reduction in total paired-end read counts was evident at the boundaries of Junction 1 and 2, flanking the 143 kb CxSV segment and extending to the boundaries defined by Junction 3 (Figure 1D). This pattern of reduced read number and homozygosity for all variant calls within this boundary was interpreted as haploinsufficiency (copy number equals 1 or CN1) and was found in all of the OCA-affected probands that had the three CxSV junctions. Overall, the CN1 regions included a 1.7 kb region in intron 1 (described above, chr15:28337022–28339402), a 17.8 kb region in intron 2 (chr15:28285968–28303784), and a 23 kb region in intron 19 (chr15:28119924–28143224) (Figure 1).

Given the frequent occurrence at which we observed these three *OCA2* structural variant junctions in our initial probands (6 out of 211; 2.8%), we queried 179 additional OCA probands (who were without definitive bimolecular diagnosis after exonic sequencing of

known OCA genes) for the 143 kb CxSV and 184 kb deletion junction fragments. PCR amplification of the junctions followed by Sanger sequencing identified four additional families with the same Junction 1 – 3 sites, bringing the total to 10 unrelated families with identical junctions and associated novel sequence insertions (Supp. Figure S1). Junction fragments for both the 143 kb CxSV and 184 kb deletion segregated in *cis* among family members (Supp. Figure S2), confirming that both structural rearrangements are contained on a single allele (Figure 2B). We now refer to this allele as the "143kb;184kb CxSV." The large number of unrelated individuals (10/390) all exhibiting identical junctions suggests the 143kb;184kb CxSV was derived from a common founder.

Identification of an independent, yet related 143 kb complex variant allele

Family 11 showed evidence of a second *OCA2* CxSV that was distinct from that of Families 1–6, although the Junction 1 and 2 sequences were identical to those in the 143kb;184kb CxSV allele (Figure 1 A, C). The short-read summary profile for the Family 11 proband showed distinct differences in comparison to the 143kb;184kb CxSV; the total number of paired-end reads within the 143 kb region (chr15: 28143224–28285968) was increased by roughly a third (Figure 1D, F), and there was no evidence of the 184 kb deletion. Both the read depth summary profile (Figure 1B–D) and PCR assessment of the 184 kb deletion junction (Supp. Figure S2) confirmed the absence of the previously identified Junction 3 fragment marking the 184 kb deletion. The phase transmission of the 143 kb CxSV junctions was confirmed to be in *cis* by PCR amplification of junctions followed by Sanger sequencing of all junction fragments from the proband, both parents, and an OCA-affected sibling from Family 11 (Figure 2C, Supp. Figure S2). The presence of this allele (which we labeled the "143kb-F11 CxSV" allele) in a parent and sibling demonstrated it was not a *de novo* allele that arose in the Family 11 proband.

Consistent with the increased total read counts within the 143 kb region, the proband from Family 11 showed 227 heterozygous SNVs in this region with allele read counts at a ratio of ~66/34 rather than the 50/50 expected for a heterozygous individual (Supp. Table S4). This altered distribution reflects the presence of 3 copies of the 143 kb genomic region. We next assessed the phase of closely apposed heterozygous SNV pairs contained within the 143 kb region, because bi-allelic SNVs located in close proximity (present on the same individual reads or on paired reads) could be used to definitively partition the SNV read counts. This analysis identified 6 SNV pairs detected in three combinations (Supp. Figure S3), thus demonstrating that 3 distinct haplotypes are present across the 143 kb region. In summary, the genomic data for the Family 11 proband provide evidence of only the 143kb-F11 CxSV, independent of the 184 kb deletion present in 143kb;184kb CxSV (Figure 2C). Furthermore, the presence of three distinct allele pairs for a subset of paired SNVs within the duplicated region suggests that the Family 11 proband carries three genomic copies of this region in their genome, and the two duplicated regions in *cis* have distinct SNV compositions.

Phase assessment of coding variants in family trios with the 143kb;184kb CxSV

CCS of OCA-affected individuals in Families 1–6 and Family 11 identified four exonic SNV coding variants at *OCA2:* c.913C>T:p.Arg305Trp, c.1065G>A:p.Ala355=, c.1327G>A:p.Val443Ile and c.1551C>T:p.Cys517=. For Families 2–6 and Family 11, phase

assessment was performed for p.Arg305Trp, p.Ala355=, and p.Val443Ile variants only (Supp. Table S5), with p.Cys517C= not analyzed further. The SNV encoding p.Arg305Trp (rs1800401, MAF = 0.0459) segregated with the 143kb;184kb CxSV in all individuals, placing this allele within the inverted 143 kb sequence. In contrast, the rare p.Val443Ile variant (rs121918166, MAF =0.0043 gnomAD-genomes European) was identified in only four families (Families 3, 4, 5 and 11). p.Val443Ile has previously been identified in individuals with OCA (Lee et al. 1994a; Grønskov et al. 2009), associated with reduced eye and skin pigmentation in Scandinavian populations (Andersen et al. 2016), and found at increased frequency in familial melanoma when *CDKN2A* germline mutations were absent (Potjer et al. 2019). Evaluation of phase found that the rare p.Val443Ile segregated in *trans* to the 143kb;184kb CxSV in all 4 families (Supp. Table S5). Interestingly, the synonymous variant p.Ala355= (rs1800404) also segregated in *trans* to the 143kb;184kb CxSV allele for these four families. Therefore, the SNVs for p.Val443Ile and p.Ala355= reside in cis on the same allele, and together are located in trans to the 143kb;184kb CxSV allele (Supp. Table S5).

Deleterious *trans*-segregating variants had been previously identified by Sanger sequencing for probands in Family 7 (c.1901T>A:p.Ile634Asn) and Family 8 (c.1025A>G: p.Tyr342Cys). However, individuals in Families 9 and 10 remain without definitive molecular diagnosis for a second additional deleterious OCA allele at this time. While probands from Families 1, 2 and 6 each underwent CCS, no additional deleterious *OCA2* alleles were identified in *trans* for the 143kb;184kb CxSV. Therefore, CCS results across all 37 genes were evaluated in these 3 probands, to identify alleles that could contribute to pigmentation phenotypes in these probands (Supp. Table S7). Interestingly, the proband from Family 6 possesses multiple variants in other OCA and well-characterized pigmentation loci. Their genotype includes a rare, heterozygous SNV at *TYRP1* (c.457C>T:p.Arg153Cys), heterozygosity for an intron 4 enhancer SNV at *IRF4* (rs12203592), and homozygosity for the *TYR* c.1205G>A:p.Arg402Gln SNV (rs1126809). The presence of the last two together are of note, as rs12203592 impacts *IRF4* expression and IRF4 has been found to regulate *TYR* levels (Praetorius et al. 2013). The TYR p.Arg402Gln allele is a temperature-sensitive allele which has reduced catalytic activity and displays increased ER retention (Tripathi et al. 1991; Jagirdar et al. 2014). The presence of these variants and the *OCA2* 143kb;184kb CxSV in one individual indicates that heterozygosity for a combination of deleterious variants in multiple genes may represent a polygenic cause of OCA. Compound heterozygosity for OCA-associated variants has been previously proposed for other albinism patients (Hutton and Spritz 2008; Campbell et al. 2019).

674 SNVs are associated with the 143kb;184kb CxSV allele

Assessment of all CCS data at the *OCA2* locus identified 30/221 OCA patients carrying the hypomorphic p.Val443Ile allele. All 30 individuals carried the p.Ala355= allele, consistent with the p.Ala355= again being present in *cis* with p.Val443Ile. Genotype comparison across the *OCA2* locus among these 30 individuals expanded the p.[Val443Ile;Ala355=] allele-associated haplotype across a 46,852 bp region from intron 6 to intron 14 (94 SNVs, chr15:28218003–28264855, Supp. Table S6). This OCA patient-derived, p.

[Val443Ile;Ala355=] associated haplotype was consistent with 1 of the 2 haplotypes present in 1000 Genomes which contain p.Val443Ile, and confirms that this haplotype, while rare, has been observed previously.

The OCA patient-derived, p.[Val443Ile;Ala355=]-associated haplotype provided anchor alleles in *trans*, which were used to assess and establish corresponding haplotypes for the 143kb;184kb CxSV. These analyses facilitated the identification of a 309 SNV haplotype across the 143 kb inversion region for the 143kb;184kb CxSV (Supp. Table S6). Additional allele assessment of shared alleles from the six 143kb;184kb CxSV families across the entire *OCA2* locus and partially encompassing the *HERC2* locus (chr15: 27853970– 28566813) extended the haplotype, identifying a total of 674 SNVs associated with the 143kb;184kb CxSV (Supp. Table S6).

Two distinct haplotypes define the duplicated regions present on the 143kb-F11 CxSV allele

We also utilized the *trans* p.[Val443Ile;Ala355=] allele-associated haplotype to anchor the haplotype of the 143kb-F11 CxSV allele. To address the unique 143 kb duplication structure of this allele, we also incorporated the frequency data for heterozygous SNVs with altered frequency distribution ratios of ~66/34 identified within the 143 kb region (Supp. Table S4). The p.[Val443Ile;Ala355=]-associated haplotype for heterozygous SNVs was assigned to the *trans*-chromosome in single copy, and this information, paired with the identification of each *trans* allele as an SNV with either ~66% or ~34% frequency, allowed the two remaining alleles present in *cis* on the rearranged chromosome to be inferred (see methods). This analysis identified 74 alleles which were both (1) present on the p.[Val443Ile;Ala355=] haplotype and (2) identified as 66% alleles in the proband for Family 11. This demonstrated that the two distinct 143 kb regions that comprise the 143kb-F11 CxSV allele were derived from different haplotypes (Supp. Table S6) and is consistent with the presence of closely apposed SNV pairs with three combinations (described above and in Supp. Figure S3).

The 143 kb inverted segment is defined by a 309 SNV haplotype shared by the 143kb;184kb CxSV and 143kb-F11 CxSV alleles

As described previously, the junctions which define the 143 kb inverted segment are identical for all individuals with the 143kb;184kb CxSV and the 143kb-F11 CxSV alleles, suggesting these two alleles were derived from a common founder allele. To test this hypothesis, we compared the predicted 309 SNV haplotype of the 143 kb inverted segment for Families 1–6 with the genotype observed for Family 11. This analysis found that 308/309 of the predicted alleles are consistent with a common haplotype in the inverted region. The single allele which differs from the predicted common haplotype (rs16950699, Supp. Table S6) is present as a homozygous allele in the Family 11 proband, suggesting a single base pair alteration occurred at this SNV at some point in the evolution of these two allele structures.

Rare variants define differences between the 143kb;184kb CxSV and 143kb-F11 CxSV alleles

We next assessed rare variants across the entire *OCA2* locus, to determine if they were shared by the CCS families with the 143kb;184kb CxSV allele and the 143kb-F11 CxSV allele. Seven rare variants from *OCA2*/*HERC2* CCS-targeted regions were identified (Table 1, Supp. Table S6, Figure 3A). All proband individuals and family members with the Junction 1 and 2 fragments defining the 143 kb inverted segment possessed the rare variants rs374519281 (A>G), MAF=0.0007, located in *HERC2* intron 82, and rs139696407 (G>A), MAF=0.0025, located in *OCA2* intron 4 (Figure 3A,B). These rare alleles were not present in any of the remaining CCS patients. We also identified 4 rare SNVs, each in tight LD with each other (R2'=1.0), located in *OCA2* intron 1, and a single variant, rs150153133 (T>C), located in intron 2 (Figure 3A). Interestingly, the intron 1 and intron 2 rare variants were only found with the 143kb;184kb CxSV allele and were not present with the 143kb-F11 CxSV allele (Figure 3B). Thus, the intervening sequences located within intron 1 to intron 2 defined by these rare SNVs differ between the 143kb-F11 CxSV and the 143kb;184kb CxSV. Taken together, the shared rare variants rs374519281 and rs13969640 provide evidence of a common ancestral allele linked to an initial 143 kb inversion event, and the rare variants in introns 1 and 2 that differ between the 143kb;184kb CxSV and the 143kb-F11 CxSV suggest a subsequent recombination event occurred to account for this discrepancy (Figure 4).

Haplotypes for the 143kb;184kb CxSV and 143kb-F11 CxSV alleles are not present in 1000 Genomes

Given the frequency at which we observed the *OCA2* 143kb;184kb CxSV in our patient population (10/390 probands; 2.6%), we next assessed the frequency at which the 143kb;184kb CxSV and the 143kb-F11 CxSV occur among individuals that do not exhibit albinism. For this we selected the rare intron 4 variant rs139696407, as it is present in all 11 families, fixed in the inverted region haplotype (Supp. Table S6), and its rare minor allele frequency allowed it to serve as a proxy SNV for screening populations. The SNV rs139696407 is present in 1000 Genomes with a frequency of 0.0026 (13 in 5008 alleles, LDlink) (Machiela and Chanock 2015). Closer examination of the haplotypes defined by rs139696407 found that each was distinct from the two 143kb;184kb CxSV and 143kb-F11 CxSV haplotypes (Table 1, Supp. Table S6). Thus, the 143kb;184kb CxSV and 143kb-F11 CxSV haplotypes are not observed in 1000 Genomes.

Modelling the evolutionary history of the complex *OCA2* rearrangements

We propose the following model to explain the presence of these two complex *OCA2* alleles (Figure 4). A common ancestral allele linking the two is supported by the following: (1) all families exhibit identical, novel 143 kb junctions related to the 143 kb CxSV event and (2) all families share the rare variants rs139696407, located within intron 4 of the inverted segment, and rs374519281, a SNV 36 kb proximal to the rearrangement and located in intron 82 of *HERC2* (Table 1, Figure 4). While we cannot ascertain if the ancestral 143 kb CxSV allele was initially generated by a *cis*- or *trans*-derived event, it is clear that this event preceded the additional alterations giving rise to the alleles segregating in families,

because the locations of the 184 kb deletion junctions flank the region of the 143 kb inverted segment.

This model incorporates the distinct haplotypes identified for the two 143 kb regions on the 143kb-F11 CxSV allele. Of note are rs139696407 and rs1800401 (p.Arg305Trp), both of which reside only within the 143 kb inverted segment. This model also takes into account the five rare variants located distal to the 143 kb CxSV within intron 1 and intron 2, that are present on the 143kb;184kb CxSV but not on the 143kb-F11 CxSV. This difference can be most directly explained by a single recombination event having occurred at some point in time on the common intermediate ancestral allele. Taken together, these data provide evidence of a single founder ancestral allele harboring the 143 kb CxSV event that subsequently gave rise to both the 143kb-F11 CxSV allele, through recombination, and separately to the 143kb;184kb CxSV allele, through an additional 184 kb deletion.

Discussion

The targeted sequencing of exons has been a powerful tool in identifying coding variants associated with human disease. However, this approach is not sufficient to characterize non-coding alterations that affect the levels or structure of mRNA. For example, in patients exhibiting ocular and cutaneous phenotypes consistent with albinism, standard exome sequencing of known candidate OCA loci is frequently performed to achieve a definitive molecular diagnosis. Unfortunately, this analysis method leaves a significant number of patients with unidentified alleles. Our hypothesis was that many of these patients may in fact harbor complex alterations at candidate OCA loci that remain undetected by standard exome sequencing. Therefore, we designed an approach to more fully characterize genomic loci that could contribute to the disease. We developed a high-depth, short read sequencing custom capture approach to assess genomic sequence variations of not only exons but also introns and non-coding regions of candidate OCA genes. Initial high depth sequence analysis of *OCA2* followed by novel junction validation screening identified 11 families with identical CxSV-associated junctions. Our haplotype analysis supports a model in which a 143 kb inverted segment of *OCA2* reinserted into intron 1, roughly 33 kb proximal to the native location, and established a common founder allele. This allele subsequently underwent an additional 184 kb deletion, giving rise to the 143kb;184kb CxSV structural rearrangement seen in 10 unrelated families, and also underwent allelic recombination, giving rise to the 143kb-F11 CxSV allele seen in a single family. Our analysis of both the 143kb;184kb CxSV and the 143kb-F11 CxSV reveals how an initial complex structural variant allele at the *OCA2* locus has continued to acquire genomic changes over time and is consistent with the previous observations that large inversions predispose loci to other subsequent genomic rearrangements (Puig et at. 2015).

The 143kb;184kb CxSV allele highlights the limitations of traditional exon-targeted sequencing in detecting complex structural variants. Although the *OCA2* locus has undergone striking, large-scale rearrangements in this CxSV, all exonic *OCA2* sequences are still present. This would lead to a copy-number neutral presentation in exome-based screening, and thus would not be flagged as a variant allele. We predict the 143kb;184kb CxSV allele to be deleterious, as the deletion of exons 3–19, along with inversion of

OCA2 sequence extending from intron 2 through 19 and its reinsertion into intron 1, would generate an mRNA transcript with novel splicing between exons 2 and 20, thus resulting in premature truncation and a functionally null *OCA2* allele (p.Ser77HisfsTer7).

The functional consequences of the 143kb-F11 CxSV allele are less clear, as this allele still retains all exons in the correct order and orientation. However, the 143kb-F11 allele removes 1.7kb of intron 1, while simultaneously inserting 143 kb of inverted *OCA2* sequence. This could decrease *OCA2* expression by disrupting intrachromosomal interactions between distal *cis*-regulatory features and the transcriptional start site (TSS). *OCA2* expression is regulated by a long-range enhancer located 21 kb upstream of the *OCA2* TSS. The SNV rs12913832 is located in this distal enhancer, has been shown through GWAS to be correlated with blue eye color in European populations, is identified as the predominant eGENE variant correlated with modulating *OCA2* expression in primary melanocyte eQTL studies, and directly impacts *OCA2* transcript levels (Eiberg et al. 2008; Sturm et al. 2008; Zhang et al. 2018). In addition, multiple lines of evidence suggest a regulatory function for intron 1 itself. Intron 1 contains the blue-eyed haplotype 1 (BEH1), which is associated with lighter pigmentation and defined by three SNPs that together form a single LD block: rs7495174, rs6497268/rs4778241, and rs11855019/rs4778138 (Duffy et al. 2007; Donnelly et al. 2012). The 1.7 kb deleted region is located entirely within the ~8.7 kb BEH1 region (variant rs4778241 is contained within the 1.7 kb deletion), thus this region is directly altered in both CxSV alleles. Two of the BEH1 variants are associated by GWAS with eye color, hair color, and melanoma (Han et al. 2008; Liu et al. 2015; Ransohoff et al. 2017; Galván-Femenía et al. 2018). Also, rs7495174 and rs4778241 are *OCA2* eQTLs, correlating these variants with altered levels of *OCA2* expression (Zhang et al. 2018). Examination of the ReMap human atlas using the UCSC genome browser (http://remap.univ-amu.fr/genome_tracks_page) showed two transcription factor binding peaks within BEH1, indicating potential regulatory regions. Alternatively, the BEHI region variants may be proxy SNPs that are tightly linked to a regulatory region in LD. Finally, we are unable to exclude the possibility that any isoform generated from either the 143kb;184kb CxSV or the 143kb-F11 CxSV allele may result in alternative cryptic splicing. Should any novel isoforms include inverted segments corresponding to coding exons, this could have the potential to further modulate *OCA2* expression in *trans* via an RNAi-mediated mechanism. Future examination of transcripts generated from both CxSV alleles in the appropriate tissue context would be useful to validate the effects of these variants.

Given the frequent occurrence of the 143kb;184kb CxSV *OCA2* allele among our proband cohort, we examined published data describing large deletions/rearrangements at *OCA2* for evidence of founder alleles or evidence of identical *OCA2* 143kb;184kb CxSV junctions. The most well-documented *OCA2* deletion that is a founder allele is the 2.7 kb deletion of exon 7 (Durham-Pierre et al. 1994; Stevens et al. 1995; Puri et al. 1997; Stevens et al. 1997; Mitchell and Reigada 2008). The frequency of this allele is estimated at 25–50% of *OCA2* mutations in African American populations (Durham-Pierre et al. 1994; Lee et al. 1994b) and ~80% of mutant alleles in South Africa (Stevens et al. 1995; Stevens et al. 1997). A 122.5 kb deletion, extending from intron 9 to intron 19 has been identified as a founder allele in the Navajo population (Yi et al. 2003). The 122.5 kb deletion is estimated to have occurred between 400 and 1000 years ago, with an estimated allele prevalence of

1/ 1500 to 1/ 2000 among the Navajo. For the 143kb;184kb and 143kb-F11 CxSV alleles, the identification of the intron 4 rare variant rs139696407 (located within intron 4 of the 143 kb inverted segment) is critical to our ability to characterize the ancestral allele, as it can be used a proxy for assessing prevalence among distinct populations. In GnomAD, rs139696407 has a MAF=0.0022, similar to the MAF of 0.0026 in 1000 Genomes. The rare rs139696407 variant is predominately in European individuals (Table 1), suggesting that these new CxSV alleles would be most frequent among individuals of European descent, an observation consistent with the self-reported ethnicity for individuals with both *OCA2* CxSVs.

Junctions 1 and 3 of our CxSV alleles indicated two breakpoints in *OCA2* intron 19. Interestingly, sequences within intron 19 of *OCA2* appear prone to structural rearrangement, as multiple other albinism-associated *OCA2* deletions involving breakpoints in intron 19 have been previously reported (Yi et al. 2003; Rooryck et al. 2011; Morice-Picard et al. 2014; Shahzad et al. 2017; Lasseaux et al. 2018; Gul et al. 2019; Chuan et al. 2021). In some cases, precise sequence breakpoint coordinates were not reported for intron 19-documented rearrangements (Shahzad et al. 2017; Chuan et al. 2021), thus not allowing for a direct comparison of the intron 19 junctions. In the remaining 5 studies reporting sequence breakpoint information on intron 19 rearrangement (Yi et al. 2003; Rooryck et al. 2011; Morice-Picard et al. 2014; Lasseaux et al. 2018; Gul et al. 2019), we found similarities between our complex rearrangement and *OCA2* intron 19 rearrangements in 3 patients first identified with altered copy number alleles by array comparative genomic hybridization (CGH) (Rooryck et al. 2011), and in a follow-up study, 2 additional patients exhibiting evidence of a complex rearrangement at *OCA2* (Morice-Picard et al. 2014). This analysis identified an identical Junction 3 fragment, indicating a 184 kb deletion from intron 2—intron 19 (chr15:28119923–28303782), and also found evidence of the Junction 1/RR breakpoint, but the corresponding Junction 2/LL breakpoint was not identified (Morice-Picard et al. 2014). Although full sequence resolution corresponding to Junction 2 was incomplete for this previous study, the presence of junctions that are identical to our CxSV alleles suggests this previously analyzed patient cohort may also contain individuals harboring the 143kb;184kb CxSV. Our ability to take advantage of high-depth sequencing of the extended *OCA2* locus in multiple individuals allowed us to fully identify all junction breakpoints for this complex allele and resolve the mechanism that gave rise to this allele. Additionally, a patient reported in Rooryck et al., 2011 identified by array CGH and qPCR exhibits a duplication that spans a region that is similar to the 143 kb region; therefore, it is interesting to speculate that this patient may harbor the 143kb-F11 CxSV allele or the intermediate allele (Figure 4). Overall, these studies indicate that additional OCA patients with these complex alleles may be present in other patient cohorts.

Currently, genomic tools are rapidly evolving and are being utilized to establish definitive diagnosis for more patients with OCA and other rare disorders. Multiple technologies have been utilized in the last few years for screening large numbers of OCA individuals, including array CGH, exome sequencing, high density custom capture-based sequencing, WGS and multiplex ligation-dependent probe amplification (MLPA) each of which have unique limitations (Alkan et al. 2011; Belkadi et al. 2015; Salpietro et al. 2018). Our focused CCS approach was successful in identifying this 143kb;184kb CxSV at *OCA2* and fully

resolving sequence junctions. The high density of short-read sequences obtained through this CCS approach also allowed full haplotype resolution for both the p.[Val443Ile;Ala355=] and 143kb;184kb CxSV alleles. However, high depth short-read sequencing does not ensure definitive diagnosis for all, and a diagnostic gap remains for individuals with OCA, similar to the diagnostic gap that is present across many other Mendelian disorders. Long-range sequencing that is independent of the current reference genome may be needed to query for large unidentified structural rearrangements of unique chromosomal architecture (Ebert et al. 2021). Additionally, we cannot exclude the potential that a subset of OCA individuals may have additional alterations in other loci that, in combination, clinically present with OCA.

Our analysis adds to the emerging body of research on structural variants and their contribution to common and rare diseases (Weischenfeldt et al. 2013, Giner-Delgado et al. 2019), as we have identified structural variants at *OCA2* that account for a portion of the missing heritability observed in albinism. Importantly, our data identifies rare variants that are in *cis* to the CxSV allele that can be incorporated into preliminary screening for the haplotypes upon which these variants arose in addition to providing unique, PCR-based assays for these junctions. Given the frequency at which we find these alleles in our own cohort (2.8%; 11/390), we suggest that individuals without definitive diagnosis by other methodologies should be directly screened for these CxSV alleles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank the families with OCA who have contributed to this study. This research was supported by the Intramural Research Program of the National Institutes of Health (NIH) and was funded by the National Human Genome Research Institute (NHGRI: 1ZIAHG000136-2 and NHGRI: 1ZIAHG000215-18). All authors have no conflicts of interest to declare.

Data Availability Statement

The deidentified patient data for the pathological structural variants reported here have been deposited to Clinvar under accession numbers SCV001623011 and SCV001623012, corresponding to the 143kb;184kb CxSV allele and the 143kb-F11 CxSV allele, respectively. Data from non-deidentified patients are available from the authors if planned research aligns with consent documents.

References

Adhikari K, Mendoza-Revilla J, Sohail A, Fuentes-Guajardo M, Lampert J, Chacón-Duque JC, Hurtado M, Villegas V, Granja V, Acuña-Alonzo V, Jaramillo C, Arias W, et al.2019. A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. Nat Commun10: 358–16. [PubMed: 30664655]

Alkan C, Sajjadian S, Eichler EE. 2011. Limitations of next-generation genome sequence assembly. Nat Meth8: 61–65.

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. Experimental Cell Research268: 26–35. [PubMed: 11461115]

Andersen JD, Pietroni C, Johansen P, Andersen MM, Pereira V, Børsting C, Morling N. 2016. Importance of nonsynonymous OCA2 variants in human eye color prediction. Mol Genet Genomic Med4: 420–430. [PubMed: 27468418]

Anikster Y, Huizing M, Anderson PD, Fitzpatrick DL, Klar A, Gross-Kieselstein E, Berkun Y, Shazberg G, Gahl WA, Hurvitz H. 2002. Evidence that Griscelli syndrome with neurological involvement is caused by mutations in RAB27A, not MYO5A. Am J Hum Genet71: 407–414. [PubMed: 12058346]

Aquaron R1990. Oculocutaneous albinism in Cameroon. A 15-year follow-up study. Ophthalmic Paediatr Genet11: 255–263. [PubMed: 2096353]

Asgari MM, Wang W, Ioannidis NM, Itnyre J, Hoffmann T, Jorgenson E, Whittemore AS. 2016. Identification of Susceptibility Loci for Cutaneous Squamous Cell Carcinoma. J Invest Dermatol136: 930–937. [PubMed: 26829030]

Awe OO, Azeke TA. 2018. Cutaneous Cancers in Nigerian Albinos: A Review of 22 Cases. Niger J Surg24: 34–38. [PubMed: 29643732]

Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova J-L, Abel L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. P Natl Acad Sci Usa112: 5473–5478.

Bellono NW, Escobar IE, Lefkovith AJ, Marks MS, Oancea E. 2014. An intracellular anion channel critical for pigmentation. Elife3: e04543. [PubMed: 25513726]

Campbell P, Ellingford JM, Parry NRA, Fletcher T, Ramsden SC, Gale T, Hall G, Smith K, Kasperaviciute D, Thomas E, Lloyd IC, Douzgou S, et al.2019. Clinical and genetic variability in children with partial albinism. Sci Rep9: 16576–10. [PubMed: 31719542]

Chahal HS, Lin Y, Ransohoff KJ, Hinds DA, Wu W, Dai H-J, Qureshi AA, Li W-Q, Kraft P, Tang JY, Han J, Sarin KY. 2016. Genome-wide association study identifies novel susceptibility loci for cutaneous squamous cell carcinoma. Nat Commun7: 12048–8. [PubMed: 27424798]

Chuan Z, Yan Y, Hao S, Zhang Q, Zhou B, Feng X, Wang X, Liu F, Zheng L, Cao Z, Ma X. 2021. Mutation Analysis of 63 Northwest Chinese Probands with Oculocutaneous Albinism. Curr Eye Res46: 140–143. [PubMed: 32552135]

Crawford NG, Kelly DE, Hansen MEB, Beltrame MH, Fan S, Bowman SL, Jewett E, Ranciaro A, Thompson S, Lo Y, Pfeifer SP, Jensen JD, et al.2017. Loci associated with skin pigmentation identified in African populations. Science358: eaan8433. [PubMed: 29025994]

Creel DJ, Summers CG, King RA. 1990. Visual anomalies associated with albinism. Ophthalmic Paediatr Genet11: 193–200. [PubMed: 2280977]

Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Barta C, Lu R-B, Zhukova OV, Kim J-J, Siniscalco M, New M, Li H, Kajuna SLB, et al.2012. A global view of the OCA2-HERC2 region and pigmentation. Hum Genet131: 683–696. [PubMed: 22065085]

Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, Hayward NK, 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 Genet80: 241–252. [PubMed: 17236130]

Durham-Pierre D, Gardner JM, Nakatsu Y, King RA, Francke U, Ching A, Aquaron R, del Marmol V, Brilliant MH. 1994. African origin of an intragenic deletion of the human P gene in tyrosinase positive oculocutaneous albinism. Nat Genet7: 176–179. [PubMed: 7920637]

Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, Yilmaz F, Zhao X, et al.2021. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science

Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L. 2008. 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. Hum Genet123: 177–187. [PubMed: 18172690]

Galván-Femenía I, Obón-Santacana M, Piñeyro D, Guindo-Martinez M, Duran X, Carreras A, Pluvinet R, Velasco J, Ramos L, Aussó S, Mercader JM, Puig L, et al.2018. Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. J Med Genet55: 765–778. [PubMed: 30166351]

Giner-Delgado C, Villatoro S, Lerga-Jaso J, Gayà-Vidal M, Oliva M, Castellano D, Pantano L, Bitarello BD, Izquierdo D, Noguera I, Olalde I, Delprat A, et al.2019. Evolutionary and functional impact of common polymorphic inversions in the human genome. Nat Commun10: 4222–14. [PubMed: 31530810]

Grønskov K, Ek J, Sand A, Scheller R, Bygum A, Brixen K, Brondum-Nielsen K, Rosenberg T. 2009. Birth prevalence and mutation spectrum in danish patients with autosomal recessive albinism. Invest Ophthalmol Vis Sci50: 1058–1064. [PubMed: 19060277]

Grønskov K, Jespersgaard C, Bruun GH, Harris P, Brondum-Nielsen K, Andresen BS, Rosenberg T. 2019. A pathogenic haplotype, common in Europeans, causes autosomal recessive albinism and uncovers missing heritability in OCA1. Sci Rep9: 645. [PubMed: 30679655]

Gul H, Shah AH, Harripaul R, Mikhailov A, Prajapati K, Khan E, Ullah F, Zubair M, Ali MZ, Shah AH, Salman S, Khan S, et al.2019. Genetic studies of multiple consanguineous Pakistani families segregating oculocutaneous albinism identified novel and reported mutations. Annals of Human Genetics83: 278–284. [PubMed: 30868578]

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, et al.2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. Plos Genet4: e1000074. [PubMed: 18483556]

Hawkes JE, Cassidy PB, Manga P, Boissy RE, Goldgar D, Cannon-Albright L, Florell SR, Leachman SA. 2013. Report of a novel OCA2 gene mutation and an investigation of OCA2 variants on melanoma risk in a familial melanoma pedigree. J Dermatol Sci69: 30–37. [PubMed: 23103111]

Hutton SM, Spritz RA. 2008. A comprehensive genetic study of autosomal recessive ocular albinism in Caucasian patients. Invest Ophthalmol Vis Sci49: 868–872. [PubMed: 18326704]

Jagirdar K, Smit DJ, Ainger SA, Lee KJ, Brown DL, Chapman B, Zhen Zhao Z, Montgomery GW, Martin NG, Stow JL, Duffy DL, Sturm RA. 2014. Molecular analysis of common polymorphisms within the human Tyrosinase locus and genetic association with pigmentation traits. Pigment Cell and Melanoma Research27: 552–564. [PubMed: 24739399]

Kausar T, Bhatti MA, Ali M, Shaikh RS, Ahmed ZM. 2013. OCA5, a novel locus for non-syndromic oculocutaneous albinism, maps to chromosome 4q24. Clin Genet84: 91–93. [PubMed: 23050561]

King RA, Summers CG. 1988. Albinism. Dermatol Clin6: 217–228. [PubMed: 3288382]

Kromberg JG, Castle D, Zwane EM, Jenkins T. 1989. Albinism and skin cancer in Southern Africa. Clin Genet36: 43–52. [PubMed: 2766562]

Kruijt CC, de Wit GC, Bergen AA, Florijn RJ, Schalij-Delfos NE, van Genderen MM. 2018. The Phenotypic Spectrum of Albinism. Ophthalmology125: 1953–1960. [PubMed: 30098354]

Landi MT, Bishop DT, Macgregor S, Machiela MJ, Stratigos AJ, Ghiorzo P, Brossard M, Calista D, Choi J, Fargnoli MC, Zhang T, Rodolfo M, et al.2020. Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. Nat Genet52: 494–504. [PubMed: 32341527]

Lasseaux E, Plaisant C, Michaud V, Pennamen P, Trimouille A, Gaston L, Monfermé S, Lacombe D, Rooryck C, Morice-Picard F, Arveiler B. 2018a. Molecular characterization of a series of 990 index patients with albinism. Pigment Cell and Melanoma Research31: 466–474. [PubMed: 29345414]

Law MH, Bishop DT, Lee JE, Brossard M, Martin NG, Moses EK, Song F, Barrett JH, Kumar R, Easton DF, Pharoah PDP, Swerdlow AJ, et al.2015. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat Genet47: 987–995. [PubMed: 26237428]

Le L, Escobar IE, Ho T, Lefkovith AJ, Latteri E, Haltaufderhyde KD, Dennis MK, Plowright L, Sviderskaya EV, Bennett DC, Oancea E, Marks MS. 2020. SLC45A2 protein stability and regulation of melanosome pH determine melanocyte pigmentation. Mol Biol Cell31: 2687–2702. [PubMed: 32966160]

Lee ST, Nicholls RD, Bundey S, Laxova R, Musarella M, Spritz RA. 1994a. Mutations of the P gene in oculocutaneous albinism, ocular albinism, and Prader-Willi syndrome plus albinism. N. Engl. J. Med330: 529–534. [PubMed: 8302318]

Lee ST, Nicholls RD, Schnur RE, Guida LC, Lu-Kuo J, Spinner NB, Zackai EH, Spritz RA. 1994b. Diverse mutations of the P gene among African-Americans with type II (tyrosinase-positive) oculocutaneous albinism (OCA2). Hum Mol Genet3: 2047–2051. [PubMed: 7874125]

Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, Zhong K, Walsh S, Chaitanya L, Wollstein A, Zhu G, Montgomery GW, et al.2015. Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. Hum Genet134: 823–835. [PubMed: 25963972]

Liyanage UE, Law MH, Han X, An J, Ong J-S, Gharahkhani P, Gordon S, Neale RE, Olsen CM, 23andMe Research Team, Macgregor S, Whiteman DC. 2019. Combined analysis of keratinocyte cancers identifies novel genome-wide loci. Hum Mol Genet28: 3148–3160. [PubMed: 31174203]

Luande J, Henschke CI, Mohammed N. 1985. The Tanzanian human albino skin. Natural history. Cancer55: 1823–1828. [PubMed: 3978567]

Lund PM. 1996. Distribution of oculocutaneous albinism in Zimbabwe. J Med Genet33: 641–644. [PubMed: 8863154]

Lyon MF, King TR, Gondo Y, Gardner JM, Nakatsu Y, Eicher EM, Brilliant MH. 1992. Genetic and molecular analysis of recessive alleles at the pink-eyed dilution (p) locus of the mouse. Proc Natl Acad Sci USA89: 6968–6972. [PubMed: 1495987]

Machiela MJ, Chanock SJ. 2015. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics31: 3555–3557. [PubMed: 26139635]

Ménasché G, Ho CH, Sanal O, Feldmann J, Tezcan I, Ersoy F, Houdusse A, Fischer A, de Saint Basile G. 2003. Griscelli syndrome restricted to hypopigmentation results from a melanophilin defect (GS3) or a MYO5A F-exon deletion (GS1). J Clin Invest112: 450–456. [PubMed: 12897212]

Mitchell CH, Reigada D. 2008. Purinergic signalling in the subretinal space: a role in the communication between the retina and the RPE. Purinergic Signal4: 101–107. [PubMed: 18368526]

Montoliu L, Grønskov K, Wei A-H, Martínez-García M, Fernández A, Arveiler B, Morice-Picard F, Riazuddin S, Suzuki T, Ahmed ZM, Rosenberg T, Li W. 2014a. Increasing the complexity: new genes and new types of albinism. Pigment Cell and Melanoma Research27: 11–18. [PubMed: 24066960]

Morgan MD, Pairo-Castineira E, Rawlik K, Canela-Xandri O, Rees J, Sims D, Tenesa A, Jackson IJ. 2018. Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. Nat Commun9: 5271–10. [PubMed: 30531825]

Morice-Picard F, Lasseaux E, Cailley D, Gros A, Toutain J, Plaisant C, Simon D, François S, Gilbert-Dussardier B, Kaplan J, Rooryck C, Lacombe D, et al.2014. High-resolution array-CGH in patients with oculocutaneous albinism identifies new deletions of the TYR, OCA2, and SLC45A2 genes and a complex rearrangement of the OCA2 gene. Pigment Cell and Melanoma Research27: 59–71. [PubMed: 24118800]

Nathan V, Johansson PA, Palmer JM, Howlie M, Hamilton HR, Wadt K, Jönsson G, Brooks KM, Pritchard AL, Hayward NK. 2019. Germline variants in oculocutaneous albinism genes and predisposition to familial cutaneous melanoma. Pigment Cell and Melanoma Research32: 854–863. [PubMed: 31233279]

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. 2010. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet42: 30–35. [PubMed: 19915526]

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, et al.2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature461: 272–276. [PubMed: 19684571]

Oetting WS, Brilliant MH, King RA. 1996. The clinical spectrum of albinism in humans. Mol Med Today2: 330–335. [PubMed: 8796918]

Okamura K, Suzuki T. 2020. Current landscape of Oculocutaneous Albinism in Japan. Pigment Cell and Melanoma Research

Okoro AN. 1975. Albinism in Nigeria. A clinical and social study. Br. J. Dermatol92: 485–492. [PubMed: 1174464]

Orlow SJ, Brilliant MH. 1999. The pink-eyed dilution locus controls the biogenesis of melanosomes and levels of melanosomal proteins in the eye. Exp Eye Res68: 147–154. [PubMed: 10068480]

Park S, Morya VK, Nguyen DH, Singh BK, Lee H-B, Kim E-K. 2015. Unrevealing the role of P-protein on melanosome biology and structure, using siRNA-mediated down regulation of OCA2. Mol. Cell. Biochem403: 61–71. [PubMed: 25656818]

Pastural E, Barrat FJ, Dufourcq-Lagelouse R, Certain S, Sanal O, Jabado N, Seger R, Griscelli C, Fischer A, de Saint Basile G. 1997. Griscelli disease maps to chromosome 15q21 and is associated with mutations in the myosin-Va gene. Nat Genet16: 289–292. [PubMed: 9207796]

Pennamen P, Le L, Tingaud-Sequeira A, Fiore M, Bauters A, Van Duong Béatrice N, Coste V, Bordet J-C, Plaisant C, Diallo M, Michaud V, Trimouille A, et al.2020. BLOC1S5 pathogenic variants cause a new type of Hermansky-Pudlak syndrome. Genet. Med22: 1613–1622. [PubMed: 32565547]

Potjer TP, Bollem S, Grimbergen AJEM, van Doorn R, Gruis NA, van Asperen CJ, Hes FJ, van der Stoep N. 2018. Multigene panel sequencing of established and candidate melanoma susceptibility genes in a large cohort of Dutch non-*CDKN2A/CDK4* melanoma families. Int J Cancer144:2453–2464.

Praetorius C, Grill C, Stacey SN, Metcalf AM, Gorkin DU, Robinson KC, Van Otterloo E, Kim RSQ, Bergsteinsdottir K, Ogmundsdóttir MH, Magnusdottir E, Mishra PJ, et al.2013. A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. Cell155: 1022–1033. [PubMed: 24267888]

Puig M, Casillas S, Villatoro S, Cáceres M. 2015. Human inversions and their functional consequences. Brief Funct Genomics14: 369–379. [PubMed: 25998059]

Puri N, Durbam-Pierre D, Aquaron R, Lund PM, King RA, Brilliant MH. 1997. Type 2 oculocutaneous albinism (OCA2) in Zimbabwe and Cameroon: distribution of the 2.7-kb deletion allele of the P gene. Hum Genet100: 651–656. [PubMed: 9341887]

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res44: W160–5. [PubMed: 27079975]

Ransohoff KJ, Wu W, Cho HG, Chahal HC, Lin Y, Dai H-J, Amos CI, Lee JE, Tang JY, Hinds DA, Han J, Wei Q, et al.2017. Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. Oncotarget8: 17586–17592. [PubMed: 28212542]

Rashkin SR, Graff RE, Kachuri L, Thai KK, Alexeeff SE, Blatchins MA, Cavazos TB, Corley DA, Emami NC, Hoffman JD, Jorgenson E, Kushi LH, et al.2020. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. Nat Commun11: 4423–14. [PubMed: 32887889]

Rayner JE, Duffy DL, Smit DJ, Jagirdar K, Lee KJ, De'Ambrosis B, Smithers BM, McMeniman EK, McInerney-Leo AM, Schaider H, Stark MS, Soyer HP, et al.2020. Germline and somatic albinism variants in amelanotic/hypomelanotic melanoma: Increased carriage of TYR and OCA2 variants. PLoS ONE15: e0238529. [PubMed: 32966289]

Rinchik EM, Bultman SJ, Horsthemke B, Lee ST, Strunk KM, Spritz RA, Avidano KM, Jong MT, Nicholls RD. 1993. A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. Nature361: 72–76. [PubMed: 8421497]

Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant Review with the Integrative Genomics Viewer. Cancer Res77: e31–e34. [PubMed: 29092934]

Rooryck C, Morice-Picard F, Lasseaux E, Cailley D, Dollfus H, Defoort-Dhellemme S, Duban-Bedu B, de Ravel TJL, Taïeb A, Lacombe D, Arveiler B. 2011b. High resolution mapping of OCA2 intragenic rearrangements and identification of a founder effect associated with a deletion in Polish albino patients. Hum Genet129: 199–208. [PubMed: 21085994]

Rosemblat S, Durham-Pierre D, Gardner JM, Nakatsu Y, Brilliant MH, Orlow SJ. 1994. Identification of a melanosomal membrane protein encoded by the pink-eyed dilution (type II oculocutaneous albinism) gene. Proc Natl Acad Sci USA91: 12071–12075. [PubMed: 7991586]

Rosemblat S, Sviderskaya EV, Easty DJ, Wilson A, Kwon BS, Bennett DC, Orlow SJ. 1998. Melanosomal defects in melanocytes from mice lacking expression of the pink-eyed dilution gene:

correction by culture in the presence of excess tyrosine. Experimental Cell Research239: 344–352. [PubMed: 9521852]

Salpietro V, Manole A, Efthymiou S, Houlden H. 2018. A Review of Copy Number Variants in Inherited Neuropathies. Curr Genomics19: 412–419. [PubMed: 30258273]

Sarin KY, Lin Y, Daneshjou R, Ziyatdinov A, Thorleifsson G, Rubin A, Pardo LM, Wu W, Khavari PA, Uitterlinden A, Nijsten T, Toland AE, et al.2020. Genome-wide meta-analysis identifies eight new susceptibility loci for cutaneous squamous cell carcinoma. Nat Commun11: 820–8. [PubMed: 32041948]

Shahzad M, Yousaf S, Waryah YM, Gul H, Kausar T, Tariq N, Mahmood U, Ali M, Khan MA, Waryah AM, Shaikh RS, Riazuddin S, et al.2017. Molecular outcomes, clinical consequences, and genetic diagnosis of Oculocutaneous Albinism in Pakistani population. Sci Rep7: 44185–15. [PubMed: 28266639]

Simeonov DR, Wang X, Wang C, Sergeev Y, Dolinska M, Bower M, Fischer R, Winer D, Dubrovsky G, Balog JZ, Huizing M, Hart R, et al.2013. DNA variations in oculocutaneous albinism: an updated mutation list and current outstanding issues in molecular diagnostics. Hum. Mutat34: 827–835. [PubMed: 23504663]

Sitaram A, Piccirillo R, Palmisano I, Harper DC, Dell'angelica EC, Schiaffino MV, Marks MS. 2009. Localization to mature melanosomes by virtue of cytoplasmic dileucine motifs is required for human OCA2 function. Mol Biol Cell20: 1464–1477. [PubMed: 19116314]

Stevens G, Ramsay M, Jenkins T. 1997. Oculocutaneous albinism (OCA2) in sub-Saharan Africa: distribution of the common 2.7-kb P gene deletion mutation. Hum Genet99: 523–527. [PubMed: 9099845]

Stevens G, van Beukering J, Jenkins T, Ramsay M. 1995. An intragenic deletion of the P gene is the common mutation causing tyrosinase-positive oculocutaneous albinism in southern African Negroids. Am J Hum Genet56: 586–591. [PubMed: 7887411]

Sturm RA, Duffy DL, Zhao ZZ, Leite FPN, Stark MS, Hayward NK, Martin NG, Montgomery GW. 2008. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet82: 424–431. [PubMed: 18252222]

Tripathi RK, Giebel LB, Strunk KM, Spritz RA. 1991. A polymorphism of the human tyrosinase gene is associated with temperature-sensitive enzymatic activity. Gene Expr1: 103–110. [PubMed: 1820207]

Venter PA, Christianson AL, Hutamo CM, Makhura MP, Gericke GS. 1995. Congenital anomalies in rural black South African neonates--a silent epidemic?S. Afr. Med. J85: 15–20. [PubMed: 7784908]

Visconti A, Duffy DL, Liu F, Zhu G, Wu W, Chen Y, Hysi PG, Zeng C, Sanna M, Iles MM, Kanetsky PA, Demenais F, et al.2018. Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. Nat Commun9: 1684–7. [PubMed: 29739929]

Wei A, Wang Y, Long Y, Wang Y, Guo X, Zhou Z, Zhu W, Liu J, Bian X, Lian S, Li W. 2010. A comprehensive analysis reveals mutational spectra and common alleles in Chinese patients with oculocutaneous albinism. J Invest Dermatol130: 716–724. [PubMed: 19865097]

Weischenfeldt J, Symmons O, Spitz F, Korbel JO. 2013. Phenotypic impact of genomic structural variation: insights from and for human disease. Nature Reviews Genetics14: 125–138.

Yi Z, Garrison N, Cohen-Barak O, Karafet TM, King RA, Erickson RP, Hammer MF, Brilliant MH. 2003. A 122.5-kilobase deletion of the P gene underlies the high prevalence of oculocutaneous albinism type 2 in the Navajo population. Am J Hum Genet72: 62–72. [PubMed: 12469324]

Zhang M, Song F, Liang L, Nan H, Zhang J, Liu H, Wang L-E, Wei Q, Lee JE, Amos CI, Kraft P, Qureshi AA, et al.2013. Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. Hum Mol Genet22: 2948–2959. [PubMed: 23548203]

Zhang T, Choi J, Kovacs MA, Shi J, Xu M, NISC Comparative Sequencing Program, Melanoma Meta-Analysis Consortium, Goldstein AM, Trower AJ, Bishop DT, Iles MM, Duffy DL, et al.2018. Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. Genome Res28: 1621–1635. [PubMed: 30333196]

Zhong Z, Gu L, Zheng X, Ma N, Wu Z, Duan J, Zhang J, Chen J. 2019. Comprehensive analysis of spectral distribution of a large cohort of Chinese patients with non-syndromic oculocutaneous albinism facilitates genetic diagnosis. Pigment Cell and Melanoma Research32: 672–686. [PubMed: 31077556]

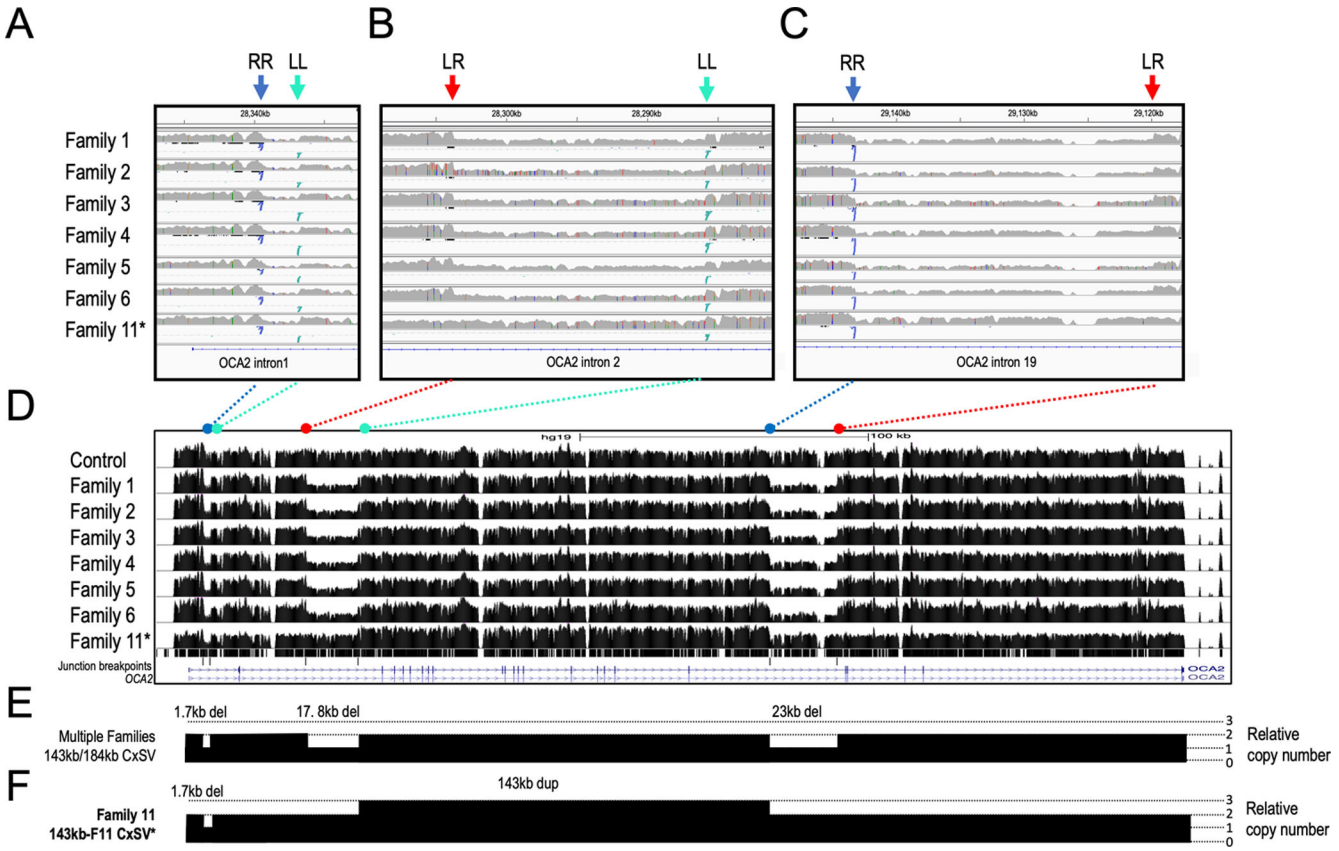

Figure 1.

Custom capture sequencing analysis identifies abnormal read pairs that define a complex *OCA2* rearrangement shared among unrelated individuals. A-C) Integrative Genomics Viewer (IGV) visualization of aberrantly mapped read pairs which highlight 3 breakpoint junctions, indicated by colored arrows and dotted lines (Junction 1/RR in dark blue, Junction 2/LL in teal blue, Junction 3/LR in red). Families 1–6 and Family 11 share aberrantly mapped read pairs for Junction 1 (intron 1 and 19, dark blue RR) and Junction 2 (intron 1 and 2, teal blue LL) which taken together denote an inversion and reinsertion of a 143 kb region (chr:28143224–2825968) into intron 1, with the simultaneous deletion of 1680 bp (chr15:28337022–28339402). Individuals in Families 1–6 also possess a 184 kb deletion, denoted by Junction 3 breakpoints (intron 2 and 19, red, LR; chr15: 28119924–28303784). D) Histograms of the read counts for each individual at the *OCA2* locus, visualized using the UCSC genome browser. Family 11 is marked by * (in D and F) to highlight that this family does not exhibit the aberrantly mapped read pairs for Junction 3 or reduced number of reads flanking the 143kb region which provide evidence of the 184 kb deletion present in Families 1–6. E-F) Graphic depiction summarizing the read counts at *OCA2* for Families 1–6 with the 143kb;184kb CxSV (E) and Family 11 with the 143kb-F11 CxSV (F), depicting how the 143 kb CxSV and 184 kb deletion events impact total copy number detected at the *OCA2* locus.

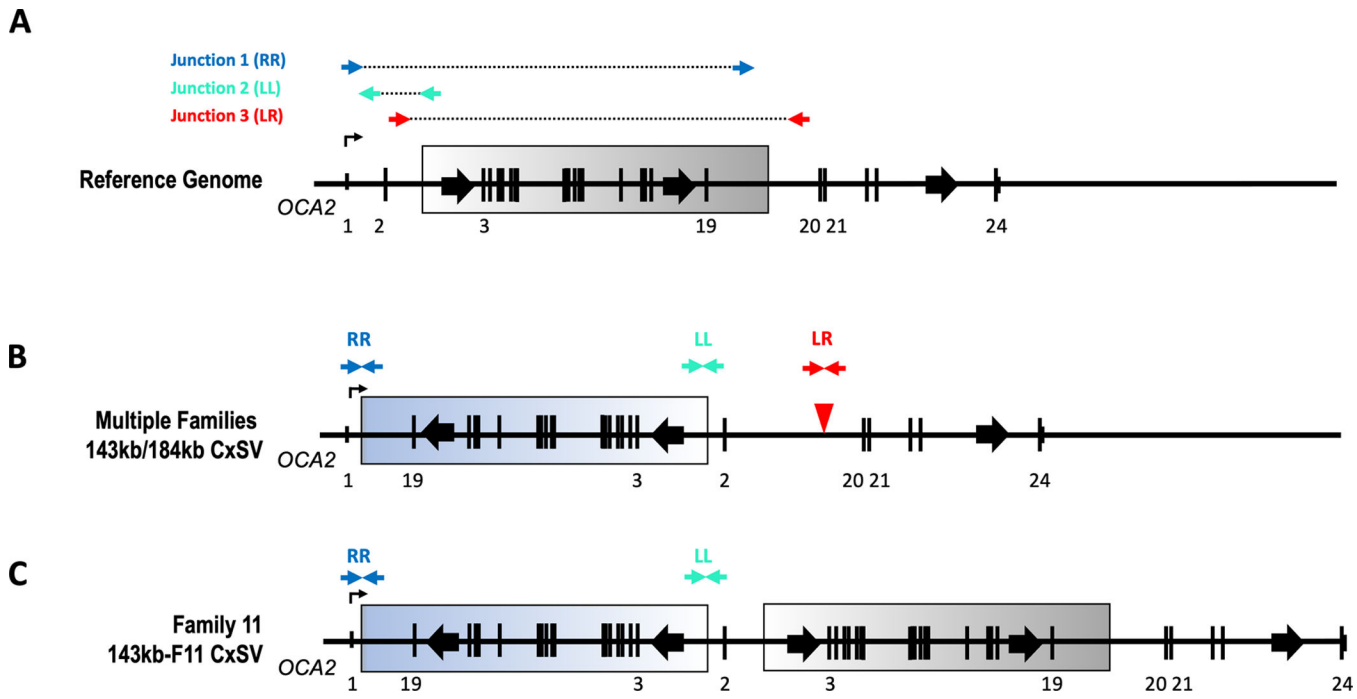

Figure 2.

Diagram of predicted chromosomal structure present in the *OCA2* 143kb;184kb CxSV allele and in the independent, yet related 143kb-F11 CxSV allele. A) The *OCA2* reference genomic locus is shown. The map locations of aberrantly mapped read pair junctions relative to the reference *OCA2* genomic locus are denoted above. B) The 143kb;184kb CxSV allele is shown, and is defined by HGVS nomenclature as NC_000015.9:g.[28337021_28339403delins[CCTGGTTGTAGGTCTAACCTGGTTAGAATCT;28143225_28285967inv; C];[28119923_28303785del]. C) The 143kb-F11 CxSV allele is shown, and is defined by HGVS nomenclature as NC_000015.9:g.28337021_28339403delins[CCTGGTTGTAGGTCTAACCTGGTTAGAATCT;28143225_28285967inv;C]. In B and C, the location and orientation of aberrantly mapped read pairs in each of the resolved CxSV alleles are indicated above each allele. (RR) indicates right/right read pair map locations defining Junction 1, and (LL) indicates left/left read pair map locations defining Junction 2. (LR) denotes the left/right orientation of aberrantly mapped Junction 3 paired-end reads. Color gradient box marks the boundary of the 143 kb inversion segment and light to dark shading marks the original 5' to 3' reference genome orientation of each region on the allele.

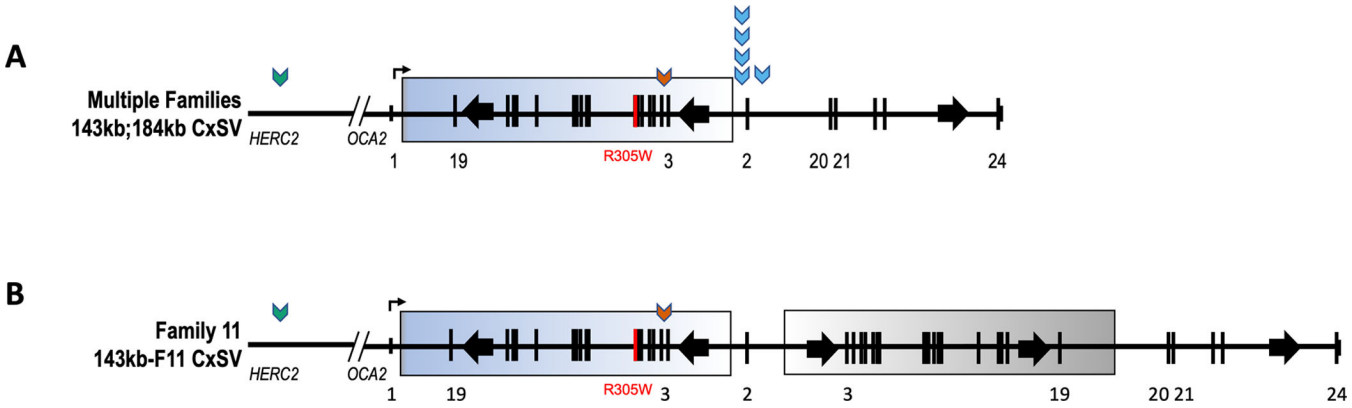

Figure 3.

Rare SNVs are associated with the complex rearrangements at *OCA2*. A) The 143kb;184kb CxSV allele, which is present in multiple families. B) The Family 11, 143kb-F11 CxSV allele. Intronic SNVs are indicated by chevrons. The orange chevron denotes the rare intron 4 SNV, rs139696407, within the 143 kb inversion-duplication region. The green chevron denotes SNV rs374519281, located outside of the 143 kb inversion-duplication region in *HERC2* intron 82. Blue chevrons denote the 5 SNVs, rs182762383, rs542133579, rs191302371 and 186994548 located in intron 1 and rs150153133 in *OCA2* intron 2, present on 143kb;184kb CxSV alleles, but not on the 143kb-F11 CxSV allele. The coding variant p.Arg305Trp, which is located in the 143 kb inverted region, is indicated in red (R305W). The color gradient boxes mark the boundaries of the 143 kb inversion segment and light to dark shading marks the original 5' to 3' reference genome orientation of each region on the allele.

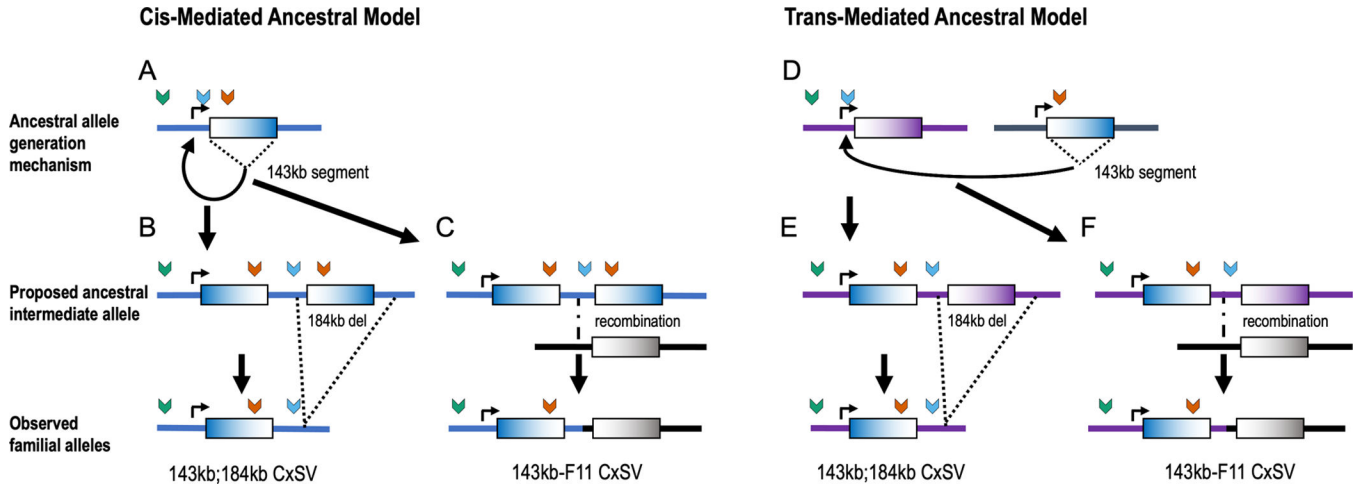

Figure 4.

Modelling the evolutionary history of the complex *OCA2* rearrangements. A *cis*-mediated ancestral model (A-C) and a *trans*-mediated ancestral model (D-F) represent the two potential mechanisms by which the 143kb;184kb CxSV allele and the 143kb-F11 CxSV allele arose. Both alleles contain identical Junctions 1 and 2, defining the 143 kb inverted segment that has been reinserted into intron 1, and the same rare variants, rs374519281 (green chevron), located 36 kb proximal to the *OCA2* TSS, and rs139696407 (orange chevron), located within the inverted sequence region. In the *cis*-mediated model (A), these two rare variants are on the same chromosome and the inversion event arose through an intra-chromosomal rearrangement. Alternatively, the two rare variants were originally located in a *trans*-orientation (D) and the inversion event was inter-chromosomal in nature. Subsequent generation of the 143kb;184kb CxSV allele is predicted by a single 184 kb deletion distal to the inversion, removing the native 143 kb region and corresponding flanking sequences (B, E). The 143kb-F11 CxSV allele, which is characterized by a duplicated 143 kb inverted region and loss of rare variants in intron 1 and 2 (blue chevron), can be accounted for by a single chromosomal recombination event in intron 1 of the proposed ancestral intermediate allele (C, F).

Table 1.

Rare variants within OCA2 intron 1 and 2 distinguish the 143kb;184kb CxSV and the 143kb CxSV-F11 alleles.

ID	Location	Alleles	143kb;184kb CxSV	143kb CxSV-F11	1K Genome alt allele frequencies, all populations	GnomAD alt allele frequencies, European	GnomAD alt allele frequencies, African	OCA cohort frequency	# OCA individuals screened
rs139696407	OCA2 Intron 4	G>A	A	A	0.0026	0.0025	0.0009	0.0282	390
rs150153133	OCA2 Intron 2	T>C	C	T	0.0032	0.0033	0.001	0.0142	211 †
rs191302371	OCA2 Intron 1	A>C	C	A	0.0032	0.0028	0.001	0.0142	211 †
rs542133579	OCA2 Intron 1	A>G	G	A	0.0032	0.0028	0.001	0.0142	211 †
rs182762383	OCA2 Intron 1	A>C	C	A	0.0032	0.0028	0.001	0.0142	211 †
rs186994548	OCA2 Intron 1	A>G	G	A	0.0034	0.0028	0.001	0.0142	211 †
rs374519281	HERC2 Intron 82	A>G	G	G	N/A	0.0004	0.0001	0.0282	390

† The 0.0142 frequencies reflect only the initial 6/211 probands identified by CCS screening that carry the rare variant.