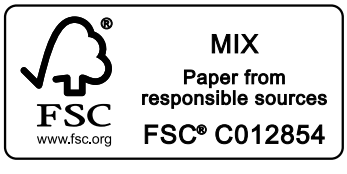
From GWAS to Function:

Transcriptional regulation of pigmentation genes in humans

Mijke Visser

ISBN/EAN: 9789490122003

The work presented in this thesis was performed at the Department of Forensic Molecular Biology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. This work was financially supported by the Erasmus MC and by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).



Layout & Cover design: Mijke Visser
Printed by: Gildeprint - Enschede

From GWAS to Function:
Transcriptional regulation of pigmentation genes in humans

Van GWAS naar functie:
Transcriptionele regulatie van pigmentatie genen in mensen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 18 maart 2015 om 11:30 uur

door
Mijke Visser
geboren te Leiden

Promotiecommissie

Promotor:	Prof.dr. M.H. Kayser
Overige leden:	Prof.dr. J. Gribnau Prof.dr. A.G. Uitterlinden Dr. R.A. Poot
Copromotor:	Dr. R.J.T.S. Palstra

Table of contents

List of abbreviations

3C	Chromosome conformation capture
ACH	Active chromatin hub
ATP	Adenosine triphosphate
BEH	Blue-eyed haploblock
bp	base pairs
cAMP	Cyclic adenosine monophosphate
CE	Capping enzyme
ChIP	Chromatin immunoprecipitation
CNV	Copy number variation
CRM	Cis-regulatory modules
CTCF	CCCTC-binding factor
CTD	Carbonyl terminal domain
DBM	DNA-binding domain
DHS	DNase I hypersensitivity
DNA	Deoxyribonucleic acid
eQTL	expression quantitative trait locus
FAIRE	Formaldehyde-Assisted Isolation of Regulatory Elements
GSH	Glutathione
GWAS	Genome-wide association study
HEK293T	Human Embryonic Kidney cell line
HPS	Hermansky-Pudlak Syndrome
IGV	Integrative Genome Browser
kb	kilo bases
kD	Kilo Dalton
LD	Linkage disequilibrium
MAPK	Mitogen-activated protein kinases
MCF7	Breast cancer cell line
MNP	Multi nucleotide polymorphism
OCA	Oculocutaneous albinism
PCF	Pause control factor
PIC	Preinitiation complex
PKA	Protein kinase A
RNA	Ribonucleic acid
RPII	RNA polymerase II
seq	sequencing
SNP	Single nucleotide polymorphism
TBP	TATA-binding protein
TF	Transcription factor
TSS	Transcription start site
UVR	Ultraviolet radiation

Scope of this thesis

Human pigmentation is one of the most explicit visual traits, which therefore has been subject of many research studies. With the emergence of large-scale genetic association studies like GWASs, numerous SNPs have been associated with a phenotype of interest, such as human eye, hair and skin color. Many of the identified pigmentation-associated SNPs have been implemented in forensic and/or anthropological applications that are developed to predict human pigmentation traits. The work described in this thesis aims to understand the functional biology underlying several of these highly associated pigmentation SNPs.

This thesis starts with a general overview of the current knowledge on human pigmentation in **Chapter 1**, including its evolutionary history and biochemistry, the mechanisms of melanogenesis, and genetic variation of pigmentation genes. It also summarizes the essentials of transcriptional regulation and the key players involved in this complex process.

Chapters 2-5 contain the experimental work performed during the course of this PhD study. Herein I focus on the biological function of SNPs that are strongly associated with human pigmentation phenotypes. In **Chapter 2**, I describe a detailed analysis of the regulatory function of an enhancer element that contains the intronic SNP rs12913832 which is strongly associated with human skin, eye and hair color, and controls expression of the pigmentation gene *OCA2*.

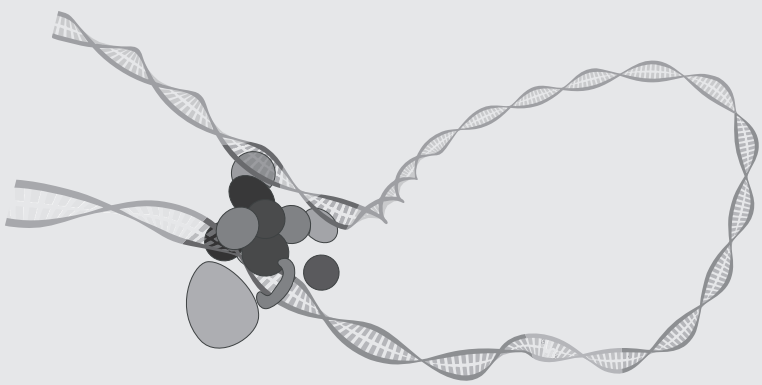
Due to the original design of GWASs and the SNP arrays used, the genetic association signals prioritized in these studies are not necessarily the actual causal or functional SNPs. These causal SNPs need to be identified in order to study the functional biology underlying the detected genetic association signals. This is exemplified in **Chapter 3**, in which I describe the identification of the actual functional SNP (rs12350739) responsible for the detected skin pigmentation-associated signal in the *BNC2* gene, followed by a detailed analysis of the transcriptional regulation of that gene and the involvement of rs12305739 therein.

Regulatory elements, such as enhancers are typically located at large distances from their target genes, however this generally does not restrict the activity of these elements, as they are able to regulate transcription over large distances through long-range interactions. **Chapter 4** focuses on chromatin structure to characterize the allele-specific regulatory mode-of-action of an intronic enhancer in which the pigmentation-associated SNP rs12203592 is located, and controls expression of the *IRF4* gene.

In **Chapter 5** I investigate the genetic basis of human skin color by combining a series of GWASs. This is followed by functional analyses of one of the five genomic regions harboring skin-color associated SNPs detected in these GWASs. At the time the work on this thesis was done, this genomic region represented the least understood genetic association signal at the functional molecular level. Finally, **Chapter 6** summarizes the results of the experimental research described in **Chapters 2-5**, and I discuss the *in silico* and experimental workflow as employed in **Chapters 2-5** to unravel the functional biology underlying genetic association signals in a more general context.

Chapter 1

General Introduction



Human pigmentation

Pigmentation is one of the most obvious visual characteristics found in living creatures. Already for centuries biologists have been intrigued with the diversity in human appearances, especially differences in skin color – being one of the most explicit phenotypes – has been subject of many research studies[1]. Skin color variation has a tremendous broad range, going from very pale to very dark, and this variation is correlated with climates, continents, and/or cultures [2]. Genetic architecture of human skin pigmentation suggests intense selective pressure in the past on important attributes of the skin, which is the largest organ of our body that most immediately and extensively interfaces with our environment[3]. This regenerative and multifunctional organ exceeds 2 m^2 in adult human, with an average thickness of only 2 mm. It forms a barrier between the body and the environment and provides protection against ultraviolet radiation (UVR), toxic reagents, pathogenic infections and mechanical stress[4]. The skin is essential for the experience of sensations like temperature, touch and pain. This is mediated through the different receptors that it contains, and provides the possibility to react appropriately to these sensations. Furthermore, the skin regulates body temperature through transpiration and is part of the Vitamin D production machinery, which is essential for stable composition of bones and teeth[5]. The skin is formed by the epidermis, which is the outer layer consisting mainly of keratinocytes; the dermis, which is a thick, dense fibro-elastic connective tissue; hair follicles; sweat glands; arteries and veins; and nerve (ending)s (Figure 1)[1].

The major hypothesis explaining the evolution of human skin color variation is that of natural selection; genetic adaptions to environmental changes as a consequence of human migration out of Africa[7]. Sunlight is suggested to be one of the most important environmental variables determining normal pigmentation variation, as a clear correlation exists between skin pigmentation and UVR[7,8], also referred to as the tanning response. In regions of high UVR, a darker skin would benefit from a better protection against sunburn and skin cancer. However, it is also suggested that variation in pigmentation evolved to regulate the UVR-penetrance, balancing the need to prevent folate photolysis but permit sufficient vitamin D photosynthesis[7]. The precise mechanisms of how UVR affects and has affected skin color variation remains subject to debate and might not be solved easily. An alternative hypothesis of sexual selection[9], which was postulated already in 1871 by Darwin[10], explains human pigmentation diversity of skin, hair and eye by reproductive variation driven through the perceived attractiveness or desirability of a particular appearance. Neither one of the hypotheses can be ruled out as player in the evolution of human skin color variation, and it is likely that a combination of the different types of selection determined the pigmentation diversity of modern humans at different stages of human evolution, and probably with different levels of impact.

Human skin color variation provides researchers from multiple disciplines, including anthropologists, geneticists and (molecular and/or cell) biologists, an ultimate opportunity to study subjects as diverse as human origins and evolution, relationships between genotypic and phenotypic diversity, and (molecular pathways determining) biogenesis and movement of subcellular organelles[2]. With increasing knowledge on DNA, transcription, translation, gene expression and regulation, research on pigmentation

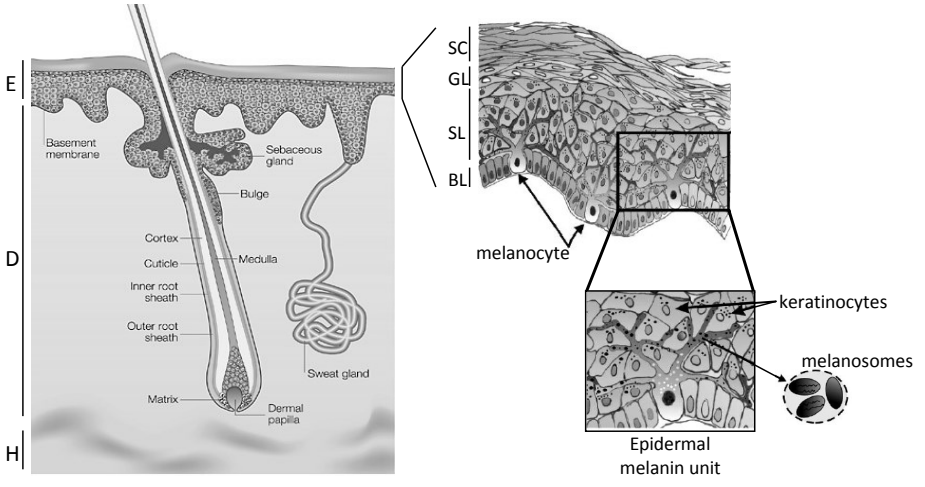


Figure 1. Schematic representation of the morphology of human skin. Left: the laminar structure of skin consists of the subcutaneous (fat) tissue (referred to as the hypodermis(H)), the dermis (D) and the epidermis (E). Right side: The epidermis is built from four different layers: basal layer (BL), spinous layer (SL), granular layer (GL) and cornified layer ('stratum corneum' (SC)). Melanocytes residing between the basal layer cells communicate through dendritic processes with 30-40 keratinocytes in the epidermal pigment unit. Melanin is synthesized by melanocytes in melanosomes, that are transferred to keratinocytes. (Adapted from Fuchs et al[4] and Cichorek et al[6]).

becomes more and more detailed in deciphering the complex mechanisms determining our variety of pigmentation phenotypes.

Melanogenesis

The color of the skin, hair and eye of mammals is primarily determined by the production and distribution of pigmented biopolymers known as melanins. Melanins are synthesized in melanocytes located in the basal layer of the skin epidermis, hair bulb and the iris. Within the melanocytes, the melanin production (referred to as melanogenesis) takes place in lysosome-like organelles, called melanosomes. Melanogenesis was first described by Seiji et al in 1961[11], after which it has been and still is intensively studied, and is the subject of several excellent reviews[1,12–14]. In skin, the melanosomes are transferred from the melanocytes to the adjacent keratinocytes that differentiate and migrate to the upper layers of the skin epidermis[14]. In hair, the process is similar, melanin is produced exclusively in the follicular melanocytes, from where they are transferred to surrounding immature precortical keratinocytes that will differentiate and migrate to form the pigmented hair shafts[15]. Unlike the skin and hair, in which melanin is continuously produced and secreted to their surrounding keratinocytes[16], in the iris the melanosomes are found only in the cytoplasm of the melanocytes within the iris stroma[17]. Differences in eye color are the result of variable amounts and quality of the melanosomes and the type of melanin packed within these organelles[18]. Epidermal melanocytes distribute their produced melanin to approximately 40 overlying basal and suprabasal keratinocytes via their elongated dendrites

and cell-cell contact, creating a so-called pigmentary unit or epidermal melanin unit[14,19] (Figure 1). Variation in skin (and hair) pigmentation cannot be explained by the number of melanocytes present in the epidermis, as this is constant in any given area, despite the level of pigmentation. Instead, this variation is due to several other factors: the shape and distribution of the melanosomes[13,20], the signaling patterns of the pigment-receiving epidermal keratinocytes[19] and the amount and type of melanin produced[13,16,18,21–23].

Biochemistry of Melanogenesis

There are two types of melanin: the yellow-to-reddish brown pheomelanin and the less-soluble black-to-brown eumelanin[12,24,25]. The synthesis of both pheomelanin and eumelanin starts with the oxidation of L-tyrosine by tyrosinase (TYR) with dopaquinone (DQ) as the product (Figure 2). In the absence of cysteine, DQ undergoes the intramolecular addition of the amino group resulting into cyclodopa, which then rapidly undergoes a redox exchange with another DQ molecule to give Dopachrome. The other product of this redox reaction, Dopa is oxidized back into DQ by TYR. Dopachrome gradually decomposes to form mostly DHI, and to a lesser extent DHICA, the latter being catalyzed by Dopachrome Tautomerase (DCT). Finally, DHI and DHICA are oxidized to form eumelanins by TYR and Tyrosinase-related protein 1 (TYRP1), respectively. The synthesis of pheomelanin and eumelanin was initially reported by Raper in 1921[26] and Mason in 1948[27], resulting in the Raper-Mason pathway, which was intensively studied afterwards, and chemically described recently by Ito et al[28]. The quantity and quality of eumelanins, their ratio of DHI- to DHICA-derived subunits and the degree of their polymerization is strongly affected by the activities of the involved TYR-related proteins TYRP1 and DCT[29]. With cysteine present, dopaquinone rapidly reacts with the sulphydryl (SH) compound of cysteine to form two cystenyldopa (CD) isomers, 5SCD and to a lesser extent 2SCD[30]. The next step is a redox exchange between DQ and the SCD's to form CD-quinones, followed by a cyclization of the CD-quinones into ortho-quinonimine (QI). Oxidization then results into benzothiazine intermediates that are finally polymerized to pheomelanin[31].

Mixed melanogenesis

Melanocytes produce melanin in different mixtures of pheomelanin and eumelanin, depending on chemical kinetics and complex pathways. The so-called 'mixed melanogenesis' is a three-step pathway starting with the production of CDs, which is an extremely fast reaction and continues as long as cysteine is present (~1 uM). The second step is the (relatively slow) oxidation of CD to give pheomelanin, which continues as long as CD is present (~9 uM). The final step is the synthesis of eumelanin that begins only after most of the CDs (and cysteine) are depleted[28,32–34]. The ratio of eumelanin and pheomelanin is therefore determined by the availability of SH compounds in melanosomes[28,35], whereas the total amount of melanin is determined by the activity of TYR and the availability of tyrosine[36]. Besides cysteine, the major SH compound in living cells, glutathione (GSH) plays an important role in the described eu/pheomelanogenesis switch. Like cysteine, GSH reacts with dopaquinone to form CDs, and depletion of this molecule (in combination

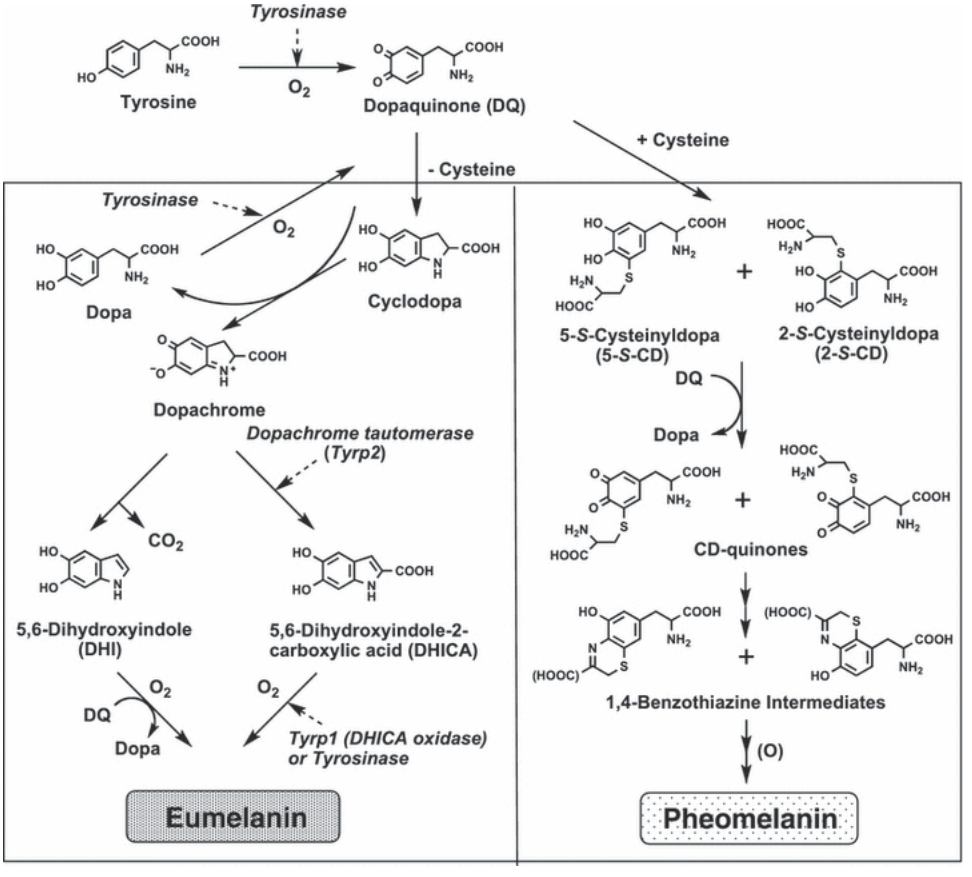


Figure 2. Biosynthetic pathways leading to eumelanin and pheomelanin synthesis. Taken from Ito et al[28].

with cysteine) results into the switch to eumelanogenesis. Additionally, both cysteine and GSH are involved in the regulation of TYR activity, however with distinct effects; cysteine inhibits the tyrosine hydroxylation reaction, whereas GSH displays an activating effect on TYR before an inhibitory effect at higher concentrations of GSH[35,37]. The subsequent increased or decreased production of DQ than leads to the branching point of eu- and pheomelanogenesis (Figure 2). In contrast to TYR, the proteins DCT (initially named TYRP2)[38] and TYRP1[39] are only involved in one side of the melanogenesis pathway, together with TYR they were shown to form a complex within melanocytes that facilitates their physiological interactions[39].

Molecular genetics of human pigmentation

Constitutive pigmentation is a polygenic trait with many different genes involved. The first genes identified to be involved in pigmentation biology were the genes encoding enzymes that regulate (parts of) the melanogenesis pathways, which are described in the previous

chapter. Research on pigmentation phenotypes either naturally occurring, disease-related or artificially induced, revealed many new pigmentation genes. How these genes determine the variation of pigmentation depends largely on the activity of the gene product, and/or on their level of expression, that needs to be regulated at a highly dedicated level[12,13]. In order to explain how this can be organized in the cell we need to go back to the very basics of the cell; to the DNA and its structural organization in the cell nucleus.

DNA – structure and content

Deoxyribonucleic acid or DNA is the carrier of genetic information in all living organisms. It is made up of the 4 'letters' - the nucleotides: G (guanine), A (adenine), T (thymine) and C (cytosine). These nucleotides selectively pair with one other nucleotide forming the complementary base pairs, G-C and A-T[40]. Usually, DNA is composed of two anti-parallel complementary polynucleotide strands assembled into a double helix structure[41]. The human genome, i.e. the entire human DNA code, is about 3 billion nucleotides long, it is divided over 46 chromosomes; two copies (one from the father, one from the mother) of 22 autosomes and 2 sex chromosomes, XX in females, XY in males.

DNA was first discovered in 1869 by Friedrich Miescher. He identified it as nuclein, as he isolated it from the cells' nuclei, a name that is preserved in today's designation deoxyribonucleic acid[42]. However, it took another 70 years before DNA was suggested to be the carrier of genetic information[43], and soon after it was found that the 4 bases are always present in fixed ratios due to their base pairing[44]. In 1952 DNA was confirmed to be the genetic material[45], after which the famous duo Francis Crick and James Watson solved the structure of DNA, supported by the X-ray analyses of Rosalind Franklin and Maurice Wilkins[41].

The double helix structure of the two intertwined nucleotide strands provides not enough compaction for the DNA to fit into the cell nucleus which has a typical diameter of only 6-10 um, as the total length of the human genome is almost 2 meters. Therefore the DNA is packed into a highly efficient and refined nucleoprotein structure known as chromatin[40]. The central subunit of chromatin is the nucleosome, which consists of approximately 165 base pairs (bp) of DNA wrapped in two superhelical turns around an octamer of core histones (assembled by two of each histones H2A, H2B, H3 and H4), resulting into a five- to tenfold compaction of DNA[46,47]. The DNA between two nucleosomes is called linker DNA, the length of this linker DNA is variable, ranging from 10 to 80 bp, and it serves different regulatory functions[48]. The stretch of nucleosomes can be folded into a compact fiber and stabilized by the binding of a fifth histone, H1, to each nucleosome and to its adjacent linker[47]. It is still poorly understood how further compaction of the chromatin is achieved to obtain higher-order chromatin, with the highest level of compaction observed during mitosis, where DNA is packed into metaphase chromosomes[47] (Figure 3). In non-dividing cells, chromatin is distributed throughout the nucleus and organized into condensed regions, called heterochromatin and into the more open and accessible regions, called euchromatin. Initially it was assumed that the purpose of chromatin was solely to package the large genome enabling it to fit in the constraints of the cell nucleus. However, now it appears that chromatin structure is at least as important for regulating

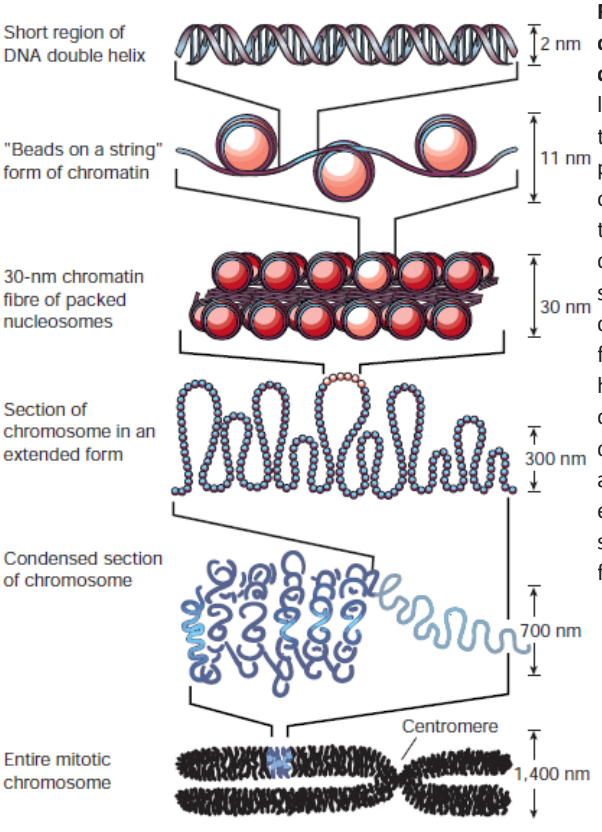


Figure 3. Packaging DNA; the organization of DNA within the chromatin structure. At the lowest level of organization, two superhelical turns of DNA (a total of 165 base pairs) are wound around the outside of a histone octamer, resulting into the nucleosome. Nucleosomes are connected to each other by linker DNA, stabilized and compacted into 30 nm chromatin fibre structures, and these fibres are then further folded into higher-order structures. The highest order of compaction is at the level of chromosomes, the details of folding at this particular level, but also at the earlier levels beyond the nucleosome structure remain uncertain. Taken from Felsenfeld et al[47].

gene expression.

The human genome can be divided into several functional units, of which the genes encode the most important information. The best-defined functional units in the human genome are the protein-coding sequences or genes, of which the number is currently estimated to be approximately 20.000[49], the majority of these genes are highly conserved throughout evolution. The synthesis of proteins starts with 'reading' the DNA-sequence of the protein-coding gene region in the genome in a process called transcription. In contrast to the replication process, in which DNA is copied into another strand of DNA, during transcription the specific region of DNA encoding the desired protein(s) or the so-called 'transcriptional unit' is copied into an RNA (ribonucleic acid) molecule. These RNA transcripts are then modified to remove the noncoding sequences (introns), resulting into the messenger (mRNA), containing only the protein-coding sequences (exons). This modification process is called splicing. Often individual genes are spliced in different ways (alternative splicing), allowing a single gene to encode multiple different protein-products. Following export of the mRNA from the nucleus to the cytoplasm, the mRNA is translated into proteins by the ribosome. Every three bases of mRNA (codons) encode an amino acid, and these amino acids are linked to form a specific protein[40]. Correctly translated, cleaved

and folded proteins fulfill most of the labor in the cell, including performing enzymatic reactions, providing building blocks for the cell matrix and regulating processes like transcription and translation. Besides proteins, noncoding RNAs (RNA molecules that are not translated into proteins) are involved in a variety of biological functions. Well-known ncRNAs are ribosomal RNA (rRNA) and tRNA which were already studied in the 1950s, but several new types of small and long ncRNAs have been identified in the last decade, and more are anticipated to follow[49,50].

Genetic variation

As described above, the DNA sequence that we are born with is basically the same throughout all our cells, and the majority of that DNA is also identical between individuals. However, the human genomes also vary quite substantially between individuals, giving rise to all kinds of (heritable) differences in human phenotypes, like physical appearances (pigmentation, body height, and many other visible traits), behavioral aspects, and disease(-susceptibility). Variation of the human genome is due to sequence variation, structural changes and DNA modifications such as methylation. Sequence variation comprises single nucleotide and multi nucleotide polymorphisms (SNPs and MNPs, respectively), small insertions or deletions (indels) of a few nucleotides. Structural variation consists of (larger) copy number variations (CNVs) like deletions, duplications and mobile-element insertions, copy-neutral inversions and translocations, and chromosomal aneuploidies[51].

For the purpose of this thesis I will focus only on sequence variations, especially that of SNPs. SNPs or point mutations are single-base-pair changes in one sequence compared to another, that occur in >1% of individuals in a sampled population. More than 38 million SNPs (of which 58% were previously unknown) were recently mapped in 14 populations worldwide by the 1000 Genomes Project[52]. Mapping and especially identifying SNPs that are associated with a certain phenotype is highly valuable for studying the biology behind those phenotypes. Genome-wide association studies (GWASs) using high-density genotyping arrays have already identified more than 11.000 SNPs that are associated with numerous physiological phenotypes[53] (http://www.genome.gov/gwastudies), and many more are anticipated to be identified in the future. GWASs investigate the nonrandom coinheritance of DNA variants (which is called the Linkage Disequilibrium, or shortly LD) by scoring the association of hundreds of thousands of SNPs with a specific phenotype[54]. Since not all known (and unknown) SNPs can be included on the genotyping chips, two selection strategies were initially used to build whole genome SNP arrays. One is to randomly select SNPs that are relatively evenly spaced across the genome, not taking into account the inter-SNP LD patterns (i.e. Affymetrix SNP arrays). The other strategy is to select tags for haplotypes (the set of SNPs that are statistically linked with each other), such as Illumina arrays[55]. Identifying the tags can lead to the identification of all polymorphic sites in the haplotypes region, including the SNP that is likely to influence a molecular or cellular process, affecting the phenotypic variation (which is called the 'causal SNP')[56]. SNPs located in exons are known as coding SNPs, as they change protein-coding DNA sequences. Non-synonymous SNPs thereby alter an amino-acid in the protein, which might affect the protein's functioning.

Initially it was thought that causative SNPs would be coding; however, it turned out that the vast majority (more than 80%) of GWAS tag SNPs located in intergenic or intronic regions. Most intergenic SNPs are located 20-50 kb away from the nearest gene in haplotype blocks of 10-20 kb, and it was estimated that in 40% of the cases these associated haplotype blocks do not overlap with known exons[57]. The effect of noncoding SNPs, intergenic as well as intronic, or the potentially associated LD SNPs, is therefore not directed at the protein functioning, instead they are more likely to influence gene regulation[54,56].

Genetic variation in pigmentation genes – coding SNPs

Numerous genome-wide and candidate-gene association studies have been employed to find genetic variations with human pigmentation phenotypes, resulting into an extensive list of pigmentation-associated DNA variants, located in exonic as well intronic and intergenic regions. I will discuss here coding as well as noncoding DNA variants in several pigmentation genes with strong effects on pigmentation phenotypes.

The coding region of the human Melanocortin-1 Receptor (*MC1R*) gene consists of only one exon and spans less than 16 kb on chromosome 16[58]. It is highly polymorphic in non-African populations with numerous allelic variants, most of which result in a single amino-acid substitution[59,60]. *MC1R* was first identified in mice on the basis of altered coat color[61,62]. The role of *MC1R* in hair pigmentation is most profound, as certain substitutions are associated with the red-hair phenotype. This phenotype also includes fair skin, inability to tan and tendency to freckle, and has been significantly linked to the development of UV-induced skin cancer[63–65], however *MC1R* was also shown to contribute to skin pigmentation characteristics in the wider non-redheaded population[66]. *MC1R* and its mutations have been extensively studied in different skin types and geographical regions, which led to different theories on the evolution of human pigmentation. Studies on evolutionary genetics suggest that in African populations the gene is under a strong functional constraint, preventing mutations to exist in *MC1R* that promote pheomelanogenesis in sub-Saharan Africa or other regions with high UV-radiation[59,67,68]. This constraint was probably lost in the populations that left the African continent, resulting into a high level of nucleotide diversity in *MC1R* leading to a range of lighter pigmentation phenotypes, depending on the severity of the *MC1R* mutation. It remains unclear whether this drive for selection was required by an enhancer need for UVR-induced vitamin D production despite a reduced protection from UVR-induced DNA damage or as a result of a yet undiscovered critical pathway[59,67].

The solute carrier family member *SLC24A5*, encoding the NCKX5 membrane transporter protein, was identified as the human homologue of a zebrafish gene that causes the 'golden' phenotype. It appears that only one SNP is responsible for the (spectrophotometrically-measured) pigmentation effect; rs1426654-A/G changing Thr111Ala, is fixed in European populations as the threonine allele rs1426654-A, whereas in African and Asian populations the ancestral alanine allele rs1426654-G is almost fixed[69]. The function of *SLC24A5* in human pigmentation is not yet known, however it was shown that in human melanocytes the activity of TYR of the rs1426654-GG and rs1426654-GA genotypes is about twofold higher than that of the rs1426654-AA genotypes[70].

Other coding SNPs and the genes they locate in, were identified to be associated

with pigmentation phenotypes by studying human pigmentation disorders, like oculocutaneous albinism OCA and related disorders such as Hermansky-Pudlak Syndrome (HSP1-8)[21,71]. There are 4 known and well-described types of OCA, ranging from OCA1A being the most severe type with a complete absence of melanin production throughout life to the milder forms of OCA1B, OCA2, OCA3 and OCA4, which display some pigment accumulation over time[72]. Recently, additional OCA types have been identified (OCA5-7) and more types may remain to be discovered[73–76]. Each of the first four OCA types (OCA1-4) can be characterized by mutations in one specific gene respectively; *TYR*, *OCA2*, *TYRP1* and *SLC45A2*/*MATP*[75,76].

The latter and the third most common form of OCA, OCA4 was initially described among the Japanese population[77], soon after it was also reported in different European countries[78]. It is caused by recessive mutations in the *SLC45A2* gene (also known as membrane-associated transporter protein (*MATP*)), another member of the solute carrier family. The molecular function of *SLC45A2* is not yet known, although the structure of the SLC45A2 protein was predicted as a 12-transmembrane spanning protein[79]; however, no substrate for transportation has been identified yet. The variant in *SLC45A2* that is most strongly associated with hair and skin pigmentation is rs16891982 encoding the protein change Leu374Phe[80]. In cultured melanocytes the rs16891982Phe/Phe-homozygotes have reduced SLC45A2 protein levels leading to mislocalization of TYR from melanosomes to the plasma membrane and incorporation of TYR into exosomes[70,81]. Another member of the solute carrier family, *SLC24A4*, has been genetically associated with eye and hair pigmentation, skin sensitivity to sun, and increased risk for cutaneous malignant melanoma[82–87]. The exact function and its role in pigmentation remains to be determined for this gene.

Mutations in the pigmentation gene *TYR* cause OCA1, the most common type of OCA among Caucasian populations. OCA1 is divided into two sub-categories: the severest form OCA1A resulting from *TYR* mutations produce no enzyme or inactive enzyme, and the less severe OCA1B resulting from mutations in *TYR* that reduce enzyme activity allowing some accumulation of melanin over time[75,88]. Mutations in *TYRP1* lead to OCA3, also known as Rufous oculocutaneous albinism[75,89]. TYRP1 is the most abundant melanosomal protein of the melanocyte, it contains 537 amino acid residues and shares 40-52 % of amino acid homology with the TYR protein[89]. Numerous SNPs in both *TYR* and *TYRP1* were shown to be genetically associated with eye, hair and skin color variation in Europeans[83,84,90–96].

The second most prevalent form of OCA worldwide is OCA2. Recessive mutations in the *OCA2* gene are the primary cause of albinism in African populations and genetic studies on albinism in Caucasians routinely detect (new) mutations in *OCA2*. Approximately 150 mutations of the *OCA2* gene are now known and new ones are still being discovered[75,97,98]. *OCA2* is located in the vicinity of the 15q11-q13 imprinted chromosomal region, which is associated with the Prader–Willi and Angelman syndromes and some patients with either one of these syndromes indeed display features of albinism[99]. The exact function of *OCA2* remains unclear although it certainly plays a central role in melanosome biogenesis and controls the eumelanin content in melanocytes[100]. The *OCA2* gene is divided into 24 coding exons spanning approximately 345 kb, the gene encodes a 110 kD protein that contains 12 transmembrane-spanning regions. The so-called P protein is thought to be involved

in small molecule transport across the melanosomal membrane[101]. As such, it regulates melanosomal pH[102] which in turn affects processing, trafficking, and activity of TYR[103,104]. For proper *OCA2* functioning, melanosomal localization is crucial; it requires a conserved consensus acidic dileucine-based sorting motif within the cytoplasmic N-terminal region of the protein[105,106]. *OCA2* was also suggested to be involved in controlling glutathione (GSH) metabolism[107], and in regulating the unfolded protein response in the endoplasmic reticulum[108,109]. Although many mutations in the *OCA2* gene are causative for pigmentation defects, several (other) polymorphisms in and around the gene have been associated with normal variation in skin, hair, and iris pigmentation[82,83,110–117]. These genetic studies indicate that the region surrounding the *OCA2* gene is one of the main genetic determinants of pigmentation differences in humans.

Genetic variation in pigmentation genes – noncoding SNPs

Several noncoding SNPs were associated with pigmentation phenotypes, such as the SNP rs12913832 in *HERC2*. This gene has no known function in pigmentation biology, but the SNP rs12913832, located in intron 86 of *HERC2,* is the most strongly associated SNP with blue eye color in European populations[111,113,117]. It is a near perfect predictor of human eye color; the C-allele is strongly associated with blue eye-color, the T-allele is strongly associated with brown-eye color[118,119]. The exact molecular mechanism of how this SNP affects human pigmentation is studied in **Chapter 2.**

The noncoding SNPs rs2153271 and rs10756819 in intron 1 of Basonuclin 2 (*BNC2*) have been genetically associated with freckles[86], and skin pigmentation in European[96] and East Asian[120] populations. Initially, *BNC2* was found to be involved in coat pigmentation in mice[121,122], and in skin pigmentation patterns in zebrafish[123,124] studies. *BNC2* encodes a DNA-binding zinc-finger protein that is highly conserved and is expressed in many human tissues, including epithelial and germ cells. Due to its very high conservation status, the function of *BNC2* is probably essential; it is most likely involved in mRNA processing like mRNA splicing, but it has also been suggested to function as a transcription factor[125,126]. In this thesis, **Chapter 3** will deal with the impact of the pigmentation-associated SNPs on the *BNC2* gene.

So far, most DNA variants associated with pigmentation were found in or close to genes that were somehow already shown to be involved in pigmentation, either in animal studies or in humans. But GWASs might also lead to the discovery of novel factors involved in the trait studied. This was the case for Interferon regulatory factor 4 (IRF4), a member of a helix-loop-helix family of DNA-binding transcription factors involved in downstream regulation of interferon signaling. IRFs are primarily associated with immune system development and response[127]. The *IRF4* gene is mainly expressed in lymphocytes, macrophages, B cells and dendritic cells, but interestingly also in melanocytic lineages[128]. The potential involvement of *IRF4* in human pigmentation was reported by Sulem et al (2007), as they identified an intergenic variant close to *IRF4* associated with freckling[83]. Notably, the expression of *IRF4* in skin melanocytes and in melanoma cell lines was described previously[129]. Further studies demonstrated that an intronic SNP rs12203592 in intron 4 of *IRF4* is strongly associated with hair, eye and skin pigmentation, tanning

response and nevus count[82,84,86,96,115]. The biological function of this SNP will be described further at the end of this chapter and studied in more detail in **Chapter 4** of this thesis.

The SNP rs12821256 is strongly associated with blond versus brown hair[83], and it is located in an intergenic region at a distance of 350 kb from the *KITLG* gene. *KITLG* encodes the c-Kit ligand that influences melanocyte proliferation and activates keratinocytes to produce promelanogenic factors[130]. Intronic SNPs located in the *LYST* gene have also been found to affect human eye color[84]. *LYST* encodes a protein that regulates intracellular protein trafficking in endosomes, and was identified as pigmentation gene in animal studies using mice[131] and cattle[132].

Agouti signaling protein (encoded by *ASIP*) is a ligand that antagonizes the function of MC1R, driving the melanin production towards the pheomelanin route[133,134]. Several DNA variants and haplotypes in the *ASIP* region at 20q11.2 were shown to be associated with eye color[135], hair color[90,135], skin pigmentation phenotypes, including skin sensitivity and freckling[90,96,136] and skin cancer [85,137–139]. The region containing the identified variants and haploblocks at 20q11.2 spans more than 20 genes, including *ASIP* but also *RALY*, *MYH7B*, *PIGU*, *GSS*, *EIFS2S* and others[140]. The exact causal DNA variant(s) and gene(s) that give rise to the pigmentation-association signals have not yet been identified for this region, which will be explored in more detail in **Chapter 5**.

How are these intronic SNPs involved in controlling transcription of the pigmentation genes they are located in? And what is the effect of the SNPs that are located in intergenic regions, more distant from the genes they might regulate? To answer these questions the complex mechanisms of controlling gene expression, also known as transcriptional regulation should be studied. The next paragraphs will deal with the basics of gene expression, transcription and the more complex mechanisms of transcriptional regulation.

Gene expression

The level of expression is tightly regulated, each gene can be expressed at different levels depending on biological function and necessity. As such, the identity of a cell is determined by the expression profile of all genes[40]. The complete set of non-specific and specific transcripts (including mRNA, ncRNA and all other types of RNA molecules), and their quantity in a cell, is referred to as the transcriptome of that cell, tissue or developmental stage/condition[141]. A set of approximately 8000 genes are ubiquitously expressed in all cells. These are the so-called house-keeping genes and are involved in general processes like metabolism, transcription, RNA processing and translation, but also in building and maintaining the cytoskeleton, cellular matrix and all kinds of transport. Another several thousand of genes are differentially expressed between divergent cell types or tissues and an even smaller set of genes is exclusively expressed in a single cell type or at a certain developmental stage[142]. For example, genes that are involved in pigmentation, like *MITF*, *TYR* and *TYRP1* display highly specific expression patterns in skin melanocytes, whereas in other cell types expression of these genes is either low or absent. Other genes were found to be almost exclusively expressed in one specific cell type or tissue and absent in others, e.g. *AMICA1*[143] and *HBB*[144] in blood, *STATH* and *HTN3*[145] in saliva and *CDSN* and *LOR*[146] in skin

epithelial cells.

With the development of hybridization-based technologies like DNA microarrays and more recently, deep-sequencing technologies like RNA-seq (RNA sequencing), whole transcriptomes can be mapped and profiled at high-throughput scales, allowing for a very precise measurement of levels of transcripts and their isoforms in all kinds of cell types, tissues and developmental stages, and from individuals with different (disease-related) phenotypes[141].

Flexibility for the cell to alter its gene expression profile in response to environmental cues is crucial for the cell's optimal functioning and survival. This is mediated via a process called signaling; through binding of signaling molecules like hormones, cytokines and growth factors to receptors at the surface of their target cell, or through cell-cell contact, the target cell is triggered to adjust gene expression profiles via a complex intracellular cascade of events[40]. A well-studied example of a signaling molecule is the hormone α-MSH affecting the MC1R protein. MC1R, a G-coupled receptor, is located on the cell surface of melanocytes and it is involved in the hormonal regulation of melanogenesis. Upon binding of agonists like α-MSH to MC1R, the adenylyl cyclase enzyme becomes activated inside the melanocytes, resulting into increased levels of intracellular cAMP (together referred to as the PKA pathway or the cAMP-dependent pathway). Up-regulation of the PKA pathway leads to induced expression levels of various genes, including that of Microphtalmia-associated transcription factor (MITF), a key player in melanocyte biology (Figure 4)[147–149]. In turn, MITF regulates expression of various pigmentation genes[150,151], including *TYR*, which ultimately drives eumelanin production[2,152,153]. Conversely, upon binding of antagonists like agouti signaling protein (ASIP) to MC1R, TYR expression decreases, resulting into a switch to pheomelanin production[133,134]. Interestingly, even without MC1R present, α-MSH is able to control melanogenesis, possibly by acting directly in melanosomes[147].

Transcription – the basics

As differential expression levels of genes are essential for cells and organisms, the control of gene expression is regulated at 6 different levels, of which the first level of transcriptional regulation plays the most important point of control for the majority of the genes[40]. Transcription is a complex process that in general proceeds in three sequential stages: initiation, elongation and termination. The initiation of eukaryotic transcription for protein-coding genes starts with the binding of the general transcription factors (TFIID, TFIIA, TFIIB, TFIIF, TFIIE and TFIIH) along with RNA polymerase II (RPII) at the promoter region of the gene to form the preinitiation complex (PIC)[40], which is facilitated by another evolutionary-conserved, multiprotein complex called Mediator[154,155]. Mediator is at least as important for transcription as the PIC; besides interacting directly with RPII, it coordinates the activity of numerous other co-factors and it is involved in both positive and negative regulation of transcription[155]. RPII is capable of unwinding DNA, synthesizing RNA and rewinding DNA; however, it is incapable of recognizing a promoter and initiating transcription, for these essential functions, the participation of the general transcription factors is required[154]. A subunit of TFIID, the TATA-binding protein (TBP) specifically binds the TATA-box DNA fragment of the promoter, resulting into the distortion of the DNA and positions the PIC

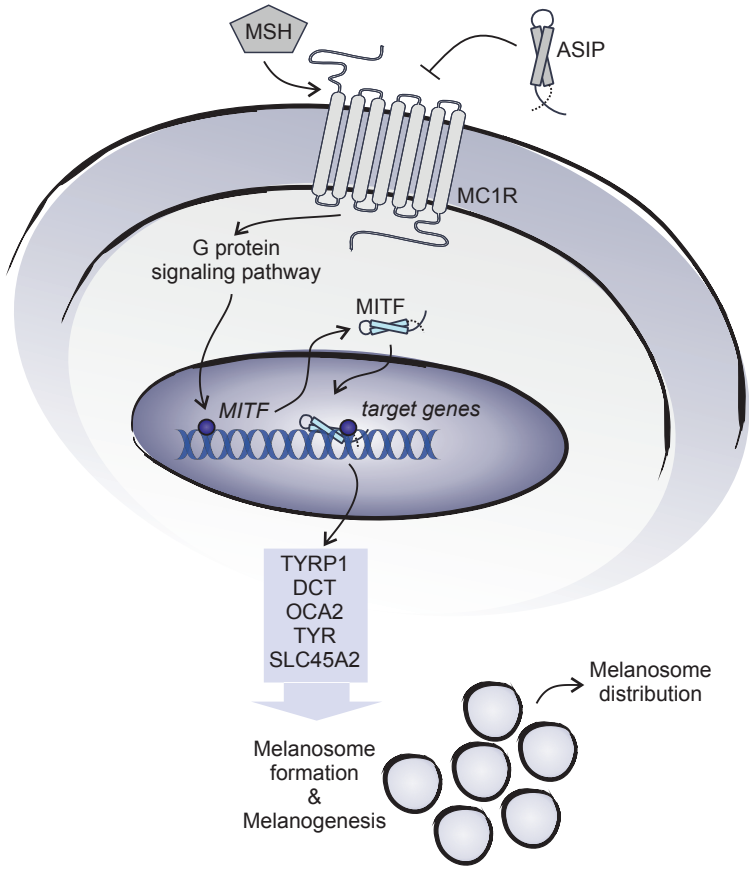


Figure 4. Melanogenesis. The trans-membrane MC1R receptor is activated upon binding of melanocyte-stimulating hormone (MSH), which leads to a cascade of signaling events; activated MC1R induces expression of the transcription factor MITF via G protein signaling. In turn, MITF regulates transcription of several genes involved in melanogenesis, including *TYRP1*, *DCT*, *OCA2*, *TYR* and *SLC45A2*, which drives eumelanin production in melanosomes, followed by distribution of these melanosomes to neighboring keratinocytes. MC1R activation can be inhibited by ASIP, which eventually leads to reduced melanogenesis and related processes. Adapted from Law et al[149].

accurately at the transcription start site (TSS)[156]. Several sequence motifs, such as the TATA-box, the TSS, the TFIIB recognition element and specific transcription factor binding sites together are called the Core Promoter Element[49]. Following this, TFIIH mediates the melting out of the DNA double helix at the TSS, allowing transcription to begin[40]. After several attempts in which RNA transcripts are released shortly after initiation resulting into truncated transcripts (a process called abortive initiation), RPII becomes phosphorylated by TFIIH, leading to the recruitment of the capping enzyme (CE) and the subsequent induction of promoter clearance. Especially the phosphorylation step is crucial for the switch from the initiation to the elongation phase; a phosphate group is added by TFIIH to the serine at position 5 (Ser5P) of the RPII carbonyl terminal domain (CTD), causing a conformational

change of RPII as such that it is released from the other factors and enters the elongation phase of transcription[157]. The RPII is paused at 20-60 bases from the TSS by pause control factors (PCFs) to ensure the complete transition of the RPII-complex to a fully competent elongating form[49,158]; however, it has also been observed that RPII is paused during later stages of elongation. Recently, emerging data indicate that pausing of RPII during early elongation is widespread and is an important step in transcriptional output regulation[158].

Progression into the elongation phase is accompanied by a decrease in Ser5P of the RPII-CTD and an increase of phosphorylation of serine 2 (Ser2P) of the RPII-CTD by the metazoan RPII Ser2 CTD kinase (P-TEFb (CycT1:Cdk9))[159]. Ser2P mediates recruitment of factors essential for the final stages of transcription to the elongating RPII (like the splicing machinery and the polyadenylation and termination machinery)[157]. In general, during elongation one DNA strand (the template strand) serves as a template to synthesize RNA. The RNPII traverses the template strand in the 5' to 3' direction at a speed of approximately 3-4 kb/min, to create an exact RNA copy, with exceptions for thymines (T) that are replaced by uracils (U) and for the nucleotides that are composed of riboses instead of deoxyriboses in the backbone. The transcription cycle ends with cleavage of the new transcript upon recognition of the termination signal, followed by multiple modifications, such as the addition of a poly-A tail (polyadenylation) at the 3'-end, capping of the 5'-end and splicing[40].

Transcriptional regulation – transcription factors, chromatin modifications and enhancers

Transcription is regulated by the combined effects of structural properties (the packaging of DNA) and the interaction of (general and specific) transcription factors on initiation and/or elongation complexes. The next paragraph will deal with both interconnected levels of transcriptional regulation (Figure 5). Transcription factors (TFs) can be separated into two classes, depending on their regulatory level; control of initiation or control of elongation[158]. However, some TFs contribute to control of both transcription stages. TFs are regulatory proteins that typically recognize and bind small 6-12bp long degenerate sequences of DNA, called motifs[160]. They are classified by the DNA-binding domains (DBMs) they possess, subdividing them into groups and families of TFs that are highly related, due to gene duplication events throughout evolution. As a result, evolutionary related TFs often share similar DBMs that in some cases leads to redundant functionality in vivo[161]. However, in many more cases, TFs with similar DBMs display highly specific functions, suggesting a high level of complexity in the regulation of TF-binding. In contrast to the general TFs that are responsible for the recruitment of the RPII complex to the genome to initiate transcription at the promoter region of all genes, (specific) TFs are more restricted to the amount of genes they transcriptionally regulate, ranging from hundreds to several thousands of genes[162]. The human genome encodes 2000-3000 TFs, many of which are expressed in a cell-type specific manner and at relatively high concentrations, as compared to the general TFs[163].

An example of a TF that is highly specific for a particular cell-type is Microphtalmia-associated transcription factor (MITF), the master transcription factor regulating melanocyte differentiation, proliferation and survival. MITF is (co)-regulating the transcription of many

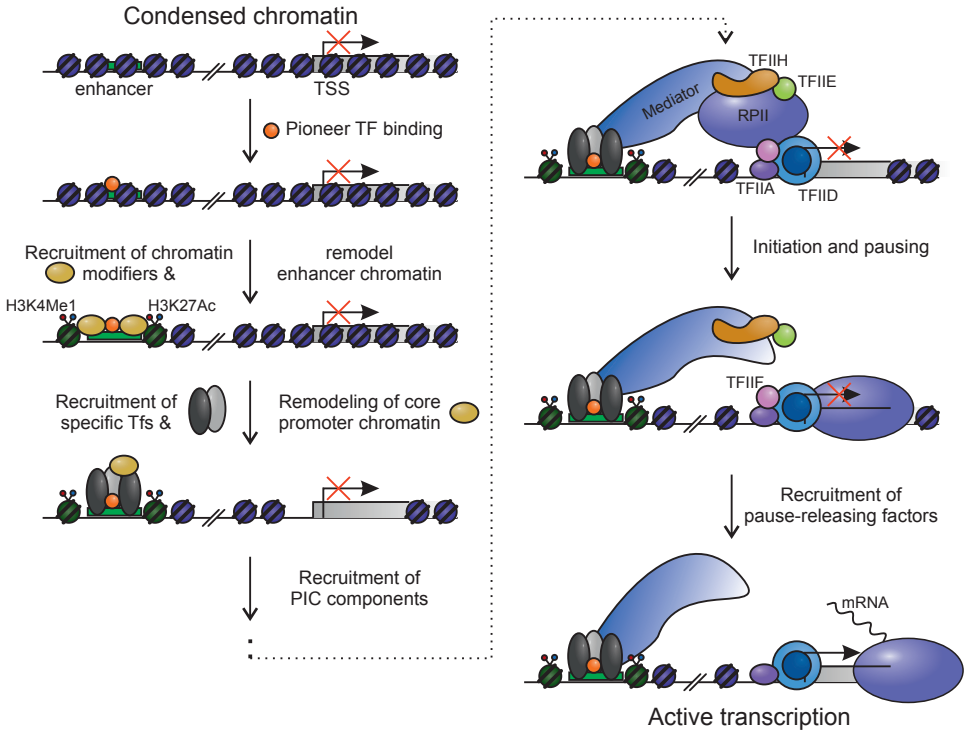


Figure 5. Transcriptional regulation and the different roles transcription factors (TFs) can play herein. Binding of pioneer TFs to an enhancer element in condensed chromatin leads to the recruitment of chromatin modifiers, that in turn remodel the enhancer chromatin. The enhancer is now accessible for cell-type or stimulus-specific TFs to bind. These TFs will then recruit factors that connect the enhancer to the core promoter, remodel the core promoter, and recruit and stabilize assembly of the preinitiation complex (PIC). Following initiation, RPII may be paused at some genes, until TFs recruit pause-releasing factors such as P-TEFb. (TSS: Transcription Start Site) Adapted from Maston et al[164].

pigmentation genes, including the melanogenic genes *TYR*, *TYRP1* and *DCT* (Figure 4), but it also targets (key) genes involved in a broad variety of other cellular processes[150,151]. Transcription of *MITF* is regulated by the TFs PAX3, SOX10, LEF1, ONECUT-2, and by MITF itself, and by the MAPK[165] and PKA signaling pathway.

Transcriptional control carried out by TFs is achieved affecting the transcription processes either directly or indirectly. The direct mechanism involves binding of TFs to the different units of the general transcription machinery like general TFs and Mediator, resulting into modulated formation of the PIC. Alternatively, TFs can interact with or recruit cofactors to factors involved in the pause-release process or elongation. The transcriptional process is indirectly affected by the induction of chromatin modifications via recruitment of chromatin modifying complexes, by competing or displacing other TFs or by acting as pioneer TFs (Figure 5), which are factors that are able to bind DNA within regions of inaccessible chromatin, creating the opportunity for other TFs to bind and activate

transcription[49,160,162,163]. Often transcription factors combine multiple strategies to regulate the transcriptional output of a gene.

The fundamental unit of chromatin; the nucleosome, also plays a central role in transcriptional regulation[48]. Mobilization of nucleosomes and the modification of histones, leads to activation or repression of transcription which depends on the specific combination of histone modifications and on DNA-accessibility. The transcription machinery and TFs are involved in the recruitment of a range of histone-modifying enzymes (called 'writers') that acetylate, methylate and ubiquitinylate or otherwise chemically modify nucleosomes across the span of each (in)active gene. With the additional presence of enzymes that remove these modifications (called 'erasers'), a highly dynamic pattern of chromatin modifications can be produced during the combined events of initiation and elongation. TFs and protein complexes contributing to transcriptional control that bind via these modifications are called 'readers'[49]. Gene activation is accompanied by recruitment of ATP-dependent chromatin remodeling complexes of the SWI/SNF family; these complexes mobilize nucleosomes to facilitate access of the transcription machinery and its regulators to DNA[166]. Gene repression, generally characterized by less accessible or fully inaccessible chromatin, can be regulated at different levels; the DNA of genes that are fully silenced is methylated and its chromatin contains different histone modifications than genes that are silenced but poised for activation at some later stage[49].

The modified amino-acids are predominantly located at the histone tails (the amino-terminal ends of the histone protein chains) of mainly Histone H3 and H4. The most frequently studied histone modifications (relevant to transcriptional regulation) are the ones at histone H3; the acetylation and tri-methylation of lysine 27 (H3K27Ac and H3K27Me3, respectively), and the mono- and tri-methylation of lysine 4 (H3K4Me1 and H3K4Me3, respectively)[47,167]. Modifications can affect one another, and for many it is known that they are mutually exclusive or are positively or negatively linked to each other. Together they define the local epigenetic status of the genetic material, which is called the histone code[168]. For example, an actively described gene will display patterns of H3K4Me3 and H3K27Ac (markers for active promoters)[169] together with tri-methylation of lysine 36 at histone3 (H3K36Me3) (a marker for actively transcribed genes), while repressed genes have patterns of H3K27Me3 (a marker of gene repression)[170] in combination with several other histone modifications (markers for heterochromatin)[171]. The histone marker H3K4Me1 is associated with enhancer elements, while the histone marker H3K27Ac distinguishes active enhancer elements from poised/inactive enhancer elements containing H3K4me1 alone[172].

Besides promoters, additional functional regulatory elements like enhancers, silencers and insulators (figure 6A), play important roles in transcription regulation. These regulatory elements are also referred to as cis-regulatory modules (CRMs). Of note, promoters are also categorized as a class of CRMs[173]. The first enhancer was identified in the early 1980s[174], since then many enhancers have been described and their functional properties have been extensively studied[167]. Recently, about one million putative enhancers were identified in the human genome by using a variety of high-throughput techniques that detect specific chromatin-associated features of enhancers[175,176]. Enhancers consist of short DNA motifs that act as binding sites for sequence-specific TFs[177], and most of the enhancers are located in DNA regions that show substantial sequence conservation.

However, relative to protein sequences, CRMs, and in particular enhancers are generally more tolerant to DNA-sequence variation, which allows for evolution to alter enhancer sequences without causing pleiotropic or lethal effects. Several studies have demonstrated that changes in regulatory DNA, like enhancers, are responsible for evolutionary adaptation in natural populations[178] and are probably the most important factors in phenotypic variation. The chromatin surrounding an (active) enhancer has a high level of specific histone modifications (as was described above) and is free of nucleosomes, it is therefore more accessible or 'open' for TF binding than nucleosome-dense regions[179]. Upon binding of the specific TFs to the enhancer, co-activators and/or co-repressors are recruited. The combination of cues coming from all factors bound to the enhancer determines the activity of the enhancer[167], and as such, modulates transcription of the target gene which can be located at a great distance on the linear chromatin template (Figure 6B)[49]. The activity of an enhancer is not restricted by the distance between the enhancer and the core promoter of its target gene. It has been shown in various studies that distant enhancer elements are brought in close proximity with the core promoter via looping[180,181]. Cell-type specific TFs

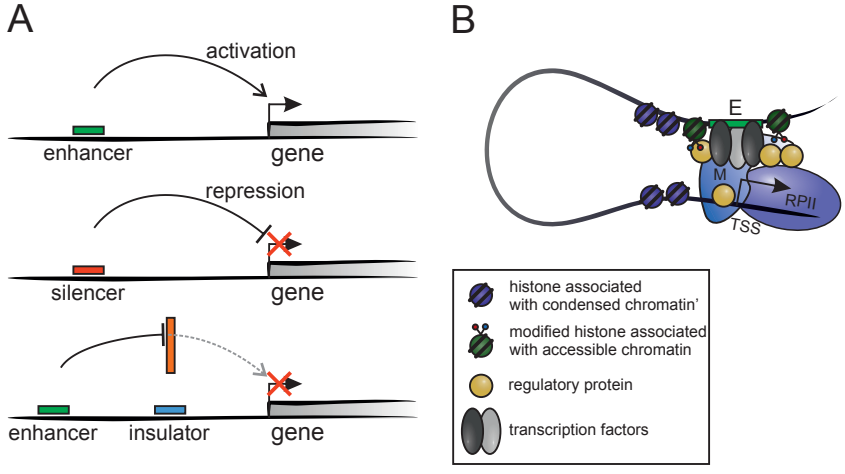


Figure 6. Functions and characteristics of cis-regulatory elements. (A) Schematic representation of the four common types of cis-regulatory elements (CRMs) and their actions. Enhancers and silencers can activate or silence their target genes from a distance. Insulators can act as barriers between CRMs and non-target genes, to prevent interactions and subsequent inappropriate regulation of these genes. Locus control regions contain combinations of (different) CRMs that cooperate to regulate transcription of a locus or gene cluster. (B) Proposed model of how an enhancer element regulates transcription by interacting with the promoter of a target gene via a long-range chromatin loop. Enhancers are usually bound by multiple transcription factors (TFs), the Mediator complex and other regulatory proteins (such as chromatin modifiers and co-activators/repressors). They are depleted of nucleosomes, while adjacent nucleosomes may contain modified histones, such as acetylated lysine 27, and mono-methylated lysine H3 of histone H3 (H3K27Ac and H3K4Me1, respectively). Looping of an enhancer to a target gene results into clustering of the recruited regulatory factors around the transcription start site (TSS) of the of the target gene to stimulate transcription.

that determine the enhancer activity are also involved in the formation (or inhibition) of enhancer-promoter loops, either directly by establishing the loop or by recruiting co-factors involved in loop formation[177]. Additional proteins involved in stabilizing chromatin loops in a genome-wide context are CCCTC-binding factor (CTCF) and cohesion, they are both crucial for large-scale organization of chromatin interactions and as such, play an important role in transcriptional regulation (for a recent review see[182]). A specific chromatin structure consisting of several chromatin loops and arranged as such that actively-transcribed genes and regulatory elements are brought together is referred to as an active chromatin hub (ACH)[183], which might be reminiscent of the transcription factory[184]. CTCF contributes to this structure by bringing different subsets of regulatory elements together[185]. The assembly of genes that are controlled by these enhancers is not random, often it involves gene-clusters that are functionally related, like the β-globin genes[186].

Enhancers are modular and flexible elements contributing additively, synergistically, and sometimes even redundantly to the overall expression pattern of their target genes[167,187]. As a consequence of these characteristics, gene-expression levels can be tightly regulated in a cell-type specific and developmental or differentiation stage-dependent manner, and can respond quickly to incoming environmental signals. An example of cell-type and gene-specific enhancer-mediated control is the transcriptional regulation of *TYR* and *TYRP1*. In mouse melanocytes, transcription of both *Tyr* and *Tyrp1* is regulated by distal regulatory elements acting on the promoter of the genes, whereas in retinal pigment epithelium the promoter alone is sufficient for expression of the genes. The distal regulatory elements of Tyr and Tyrp1 do not have similar sequence patterns, although both the enhancer elements of Tyr and Tyrp1 are activated by direct binding of Sox10, and by (probably indirect) binding of Mitf in mouse melanocytes[188], suggesting a similar mode of action.

Studying enhancer characteristics and function

Since the role of enhancers and the interaction with their target genes is an essential part of transcriptional regulation, studying existing enhancers and especially identifying novel ones remains the focus of many research efforts. There are several methods for the identification of CRMs in general, and enhancers in particular, including analysis of (clusters of) transcription factor binding motifs, mapping of epigenetic features (e.g. histone modifications and chromatin accessibility), and evolutionary conservation. Although each method separately provides valuable information, it is advisable to combine the three approaches as this will improve enhancer prediction tremendously[164,167,189,190].

Transcription factor binding motifs can be identified using computational and/or experimental approaches. Computational methods make use of either sequences that are enriched in putative regulatory regions or high average evolutionary sequence conservation of all motif occurrences[167], whereas experimental methods assay a single TF for binding at a specific region of interest or genome-wide. The technique most commonly used is *chromatin immune-precipitation* (ChIP), followed by quantitative PCR (ChIP-qPCR) or deep sequencing (ChIP-seq), for local or global views, respectively[191]. The ChIP method involves co-valently linking the TFs to their in vivo binding sites by chemical cross-linking, after which the chromatin is sheared and DNA fragments corresponding to the binding sites

are precipitated with TF-specific antibodies. Bound sequences are then detected by qPCR or deep-sequencing[192].

Analyzing evolutionary conservation can be a powerful approach for detecting regulatory elements, as a high level of sequence conservation between (related) species is indicative of purifying selection; disruptive mutations are rejected and the corresponding sequence is designated to be functional. Accelerated evolution across species or within lineages might reveal functional elements under positive selection, i.e. acquired changes or mutations in these elements increase fitness[190]. Although powerful for detecting strong, highly conserved CRMs, this approach has strong limitations especially for identifying lineage-specific, recently evolved functional elements[193].

With the development of *chromosome conformation capture* (3C) techniques it became possible to study nuclear organization at a high resolution. These methods can elucidate how chromatin fibers are folded into higher order chromatin structures by first fixing the interactions between chromatin segments that are in close proximity using formaldehyde. The cross-linked chromatin is digested using a restriction enzyme followed by intramolecular ligation of DNA fragments that were in close proximity in the nuclear space. There are several methods that are based on the 3C-technology and depending on which method is used, subsequent detection of these ligated fragments provides information about the 3D spatial organization of chromatin on the scale from single gene loci up to the whole genome[194]. In this thesis we made use of the 3C-qPCR method, in which the ligation products are detected by quantitative qPCR that measures the number of ligation events between the selected DNA fragment (the viewpoint) and the interacting fragments. This is the so-called one-to-one method in which the primers are designed near the restriction ends of the viewpoint fragment and in all the restriction fragments of interest, requiring hypotheses or expectations of potential interactions[195,196].

Furthermore, the study of epigenetic characteristics in a genome-wide fashion has been shown to be very effective in identifying functional elements. Consortia of multiple laboratories, such as the ENCyclopedia of DNA Elements (ENCODE)[197] and the US National Institutes of Health (NIH) Roadmap Epigenomics Mapping Consortium[198] are working cooperatively to map epigenetic features as histone modifications and TF-binding (using ChIP-seq), chromatin accessibility (by mapping DNaseI hypersensitivity), and expression patterns (using RNA-seq) in a wide range of cell types.

The most effective and accurate manner to predict CRMs or enhancer elements is probably a combination of the above described approaches starting with mapping epigenetic features of chromatin by analyzing histone modifications and chromatin accessibility, followed by analysis of evolutionary conservation, focusing on the regions identified in the first step. This approach should reveal the most important regulatory regions, because in alignments of orthologous sequences from a diverse set of mammals, the noncoding regions contain blocks of highly conserved sequences surrounded by blocks of low conservation. These conserved blocks can be interpreted as functional DNA sequences in which (most) substitutions were rejected during evolution[189]. Additional knowledge obtained from TF-binding analyses will improve the prediction, and will give further insight into the functionality of the identified regulatory elements (more detailed information on CRM or enhancer prediction can be found in reviews of Shlyueva et al,

Hardison et al and Kellis et al[167,189,190]). For example, a combination of different approaches was used to identify regulatory elements to produce an atlas of active enhancers. As this study was done using a large number of different human tissues and cell types, the atlas provides a valuable resource of enhancers that are active across a great diversity of tissues and cells[199]. It is however also possible to identify enhancers that are specific for one single cell type, which was shown by Gorkin et al[200]. Data obtained from ChIP-seq for EP300 (a transcriptional co-activator) and H3K4Me1 (an enhancer marker) in mouse melanocyte cells, evolutionary conservation and TF-binding motifs predicted to bind key melanocyte TFs, was combined to predict enhancer elements that are located near genes known to be involved in melanocyte biology. A subset of these predicted regulatory elements were investigated in reporter assays in melanocytes in culture and in zebrafish for confirmation, after which a 6-mers vocabulary was developed that is predictive for mouse and human melanocyte enhancer function[200].

The inadequate or absent functioning of an enhancer affects transcription of its target gene, which potentially causes or contributes to a (disease-) phenotype in humans[57,201]. Initially, variations in enhancer elements underlying specific phenotypes were identified only after the identification of the enhancer itself. But with the arisen availability of genome-wide screening methodologies, such as GWASs, phenotype-associated DNA variants could provide directions to researchers for the possible identification of novel enhancer elements[57]. The majority of these associated SNPs are not located in protein-coding regions; instead, they are located in introns or in intergenic areas, which could imply that these SNPs have a regulatory function[51]. This was shown for various SNPs associated with disease-phenotypes, such as functional, cancer risk variants at 8q24, and regulatory DNA variants associated with coronary artery disease, congenital heart disease, cardiac arrhythmias and colorectal cancer[202–210]. Generally, the associated allele disrupts a TF-binding motif, which affects the activity of the enhancer and subsequently results into altered gene expression.

Transcriptional regulation of pigmentation genes: the role of noncoding pigmentation-associated SNPs

Since numerous noncoding SNPs are associated with disease phenotypes of which several have been shown to alter an enhancer element, it can be assumed that DNA variants linked to non-disease related phenotypes, like visible traits or pigmentation variation, are also involved in transcriptional regulation. Several studies have shown that changes in regulatory DNA, like CRMs and enhancers, are responsible for evolutionary adaptation in natural populations[178] and are probably the most important factors in phenotypic variation[211,212].

At the onset of my research projects that are included in this thesis, not much was known on the potential role of noncoding SNPs in transcriptional regulation of pigmentation genes. However, a functional role had been suggested for the SNP rs12913832 that is located in intron 86 of the *HERC2* gene. *HERC2* has no known biological role in pigmentation, whereas the neighboring gene *OCA2* is a well-known pigmentation gene. The region surrounding rs12913832 is highly conserved among species, it was therefore suggested that rs12913832 is located within a distal regulatory element which modulates transcription of

OCA2 in either a negative (Figure 7a)[111] or positive (Figure 7b)[17] way. Eiberg et al reported that in colon carcinoma cells (Caco-2 cells) a 676bp *HERC2* fragment containing rs12913832 significantly reduced expression of the luciferase reporter gene in an reporter assay, and this effect was stronger with the rs12913832 T-allele present. They therefore hypothesized that the evolutionary conserved region around rs12913832 functions as a transcriptional silencer which interacts with the promoter of *OCA2*, and regulates expression of *OCA2* in an allele-specific manner[111]. In contrast, Sturm et al (2008) suggested that the evolutionary-conserved region around rs12913832 acts as an allele-dependent enhancer, rather than a silencer, regulating *OCA2* expression [117]. They proposed a model for the determination

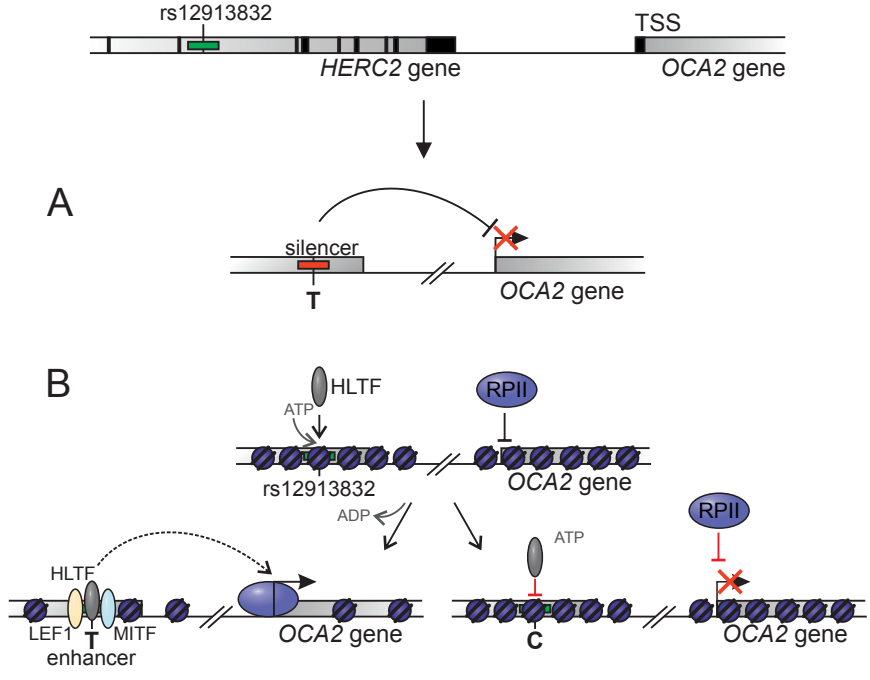


Figure 7. Proposed models for the region surrounding rs12913832 in intron 86 of *HERC2*, functioning as a regulatory element (either a silencer or an enhancer) controlling transcription of the pigmentation gene *OCA2*. (A) Model proposed by Eiberg et al, in which the region around rs12913832 acts as a long-range silencer element, which interacts with the *OCA2* promoter to control expression of *OCA2*. Activity of the proposed silencer is strongest with the rs12913832 T-allele (T) present[111]. (B) Sturm et al proposed a mechanism for the determination of blue-brown eye color in which the region around rs12913832 functions as an allele-dependent enhancer element. The transcription factor HLTF recognizes the evolutionary-conserved element containing the rs12913832 T-allele, which results into a conformational change of the chromatin to a more accessible state, allowing binding of the transcription factors MITF and LEF1. This change of chromatin state (going from heterochromatin to euchromatin) spreads to the *OCA2* promoter region, which allows for RPII to initiate transcription of *OCA2*. With the r12913832 C-allele present, HLTF, MITF and LEF1 are not able to bind, the chromatin of the *HERC2-OCA2* locus remains closed, which makes it impossible for RPII to bind, and *OCA2* is not transcribed[17].

of blue-brown eye color in which the presence of the rs12913832 T-allele allows for the transcription factor HLTF to bind the potential enhancer, followed by recruitment of the transcription factors MITF and LEF1. This results into a conformational change of the *OCA2* promoter region allowing RPII to bind and initiate transcription of *OCA2*, and the brown eye color. With the rs12913832 C-allele present, binding of HLTF and subsequent recruitment of MITF and LEF1 are prevented. As a consequence, the chromatin at the *OCA2* promoter region remains closed, which results into an absence of *OCA2* transcription and the blue eye color[17]. In **Chapter 2** both models will be explored experimentally, followed by an in-depth study of the mode-of-action of the region around rs12913832 and the transcriptional regulation of *OCA2*.

A clear example of how a pigmentation-associated DNA variant modulates expression of a nearby gene was presented recently for the *KITLG* gene and the pigmentation-associated SNP rs12821256[213]. Using sticklebacks, Miller et al were the first to give quantitative evidence for the involvement of *KITLG* in skin color variation[214]. This study demonstrated differential expression of *KITLG* between the different pigmentation phenotypes of sticklebacks, which is most likely due to mutations in regulatory elements controlling expression levels of *KITLG*, rather than coding mutations[214]. Expression of *KITLG* (Kit ligand for the Kit receptor) in skin was shown to regulate the number of melanocytes and melanin distribution during development, it is required for ongoing maintenance and survival of normal melanocyte numbers in adults[215], and is implicated in the onset of pigmentary disorders[216,217]. It was demonstrated that the intergenic SNP rs12821256 that is strongly associated with hair color variation[83], is located in an enhancer element regulating the expression of *KITLG*. This enhancer element is located about 350 kb from the TSS of *KITLG*, and the SNP rs12821256 alters the binding site for LEF1 transcription factor, thereby modulating the activity of the enhancer and subsequent expression levels of *KITLG*[213].

Furthermore, the functional biology behind the pigmentation-associated SNP rs12203592 located in intron 4 of *IRF4* has also been the subject of a recent study. This noncoding SNP was initially shown to be associated with (male-specific) childhood ALL[218]. The ALL-study provided the first evidence that rs12203592 is located within a regulatory element, controlling expression of *IRF4* through binding of transcription factor TFAP2α in an allele-specific manner. In Burkitt Lymphoma B cells, HEK293T human embryonic kidney cells and the NCI-H295R human adrenal cells the rs12203592-enhancer element was shown to bind TFAP2α with more affinity and strongly repress *IRF4*-promoter activity when the rs12203592 C-allele was present than when the T-allele was present[218]. A recent study demonstrated that in melanocytes the transcription factor TFAP2α binds to the enhancer element together with MITF, which results into activation of *IRF4* transcription[219]. However, in contrast with the results of Do et al[218], increased TFAP2α binding induced by the presence of the rs12203592 C-allele, leads to increased *IRF4*-promoter activity and subsequently to higher levels of *IRF4* expression in skin melanocytes[219]. This difference in transcriptional activation might be due to cell-type-specific effects mediated by transcription factors other than TFAP2α, such as MITF. In **Chapter 4** I studied together with my collaborators the mechanisms of transcriptional regulation of *IRF4* in more detail, especially focusing on the rs12203592 enhancer, in skin melanocytes and melanoma cell lines.

References

1. Jablonski, N. G. The Evolution of Human Skin and Skin Color. *Annu. Rev. Anthropol.* **33,** 585–623 (2004).
2. Barsh, G., Gunn, T., He, L., Schlossman, S. & Duke-Cohan, J. Biochemical and genetic studies of pigment-type switching. Pigment Cell Res. **13**, 48–53 (2000).
3. McEvoy, B., Beleza, S. & Shriver, M. D. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Hum. Mol. Genet.* **15 Spec No,** R176–81 (2006).
4. Fuchs, E. & Raghavan, S. GETTING UNDER THE SKIN OF EPIDERMAL MORPHOGENESIS. *Nat. Rev. Genet.* **3,** 199–209 (2002).
5. Scherer, D. & Kumar, R. Genetics of pigmentation in skin cancer--a review. *Mutat. Res.* **705,** 141–53 (2010).
6. Cichorek, M., Wachulska, M., Stasiewicz, A. & Tymińska, A. Skin melanocytes: biology and development. *Postępy dermatologii i Alergol.* **30,** 30–41 (2013).
7. Jablonski, N. G. & Chaplin, G. The evolution of human skin coloration. *J. Hum. Evol.* **39,** 57–106 (2000).
8. Relethford, J. H. Hemispheric difference in human skin color. *Am. J. Phys. Anthropol.* **104,** 449–57 (1997).
9. Aoki, K. Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Ann. Hum. Biol.* **29,** 589–608 (2002).
10. Darwin C. *The Descent of Man and Selection in Relation to Sex*. (J. Murray, London, 1871).
11. Seiji, M., Fitzpatrick, T. B. & Birbeck, M. S. The melanosome: a distinctive subcellular particle of mammalian melanocytes and the site of melanogenesis. *J. Invest. Dermatol.* **36,** 243–52 (1961).
12. Kondo, T. & Hearing, V. J. Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Rev. Dermatol.* **6,** 97–108 (2011).
13. Parra, E. J. Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am. J. Phys. Anthropol.* **Suppl 45,** 85–105 (2007).
14. Wu, X. & Hammer, J. A. Melanosome transfer: it is best to give and receive. *Curr. Opin. Cell Biol.* **29C,** 1–7 (2014).
15. Slominski, A. *et al.* Hair follicle pigmentation. *J. Invest. Dermatol.* **124,** 13–21 (2005).
16. Ortonne, J. P. & Prota, G. Hair melanins and hair color: ultrastructural and biochemical aspects. *J. Invest. Dermatol.* **101,** 82S–89S (1993).
17. Sturm, R. a & Larsson, M. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res.* **22,** 544–62 (2009).
18. Prota, G., Hu, D. N., Vincensi, M. R., McCormick, S. a & Napolitano, a. Characterization of melanins in human irides and cultured uveal melanocytes from eyes of different colors. *Exp. Eye Res.* **67,** 293–9 (1998).
19. Weiner, L., Fu, W., Chirico, W. J. & Brissette, J. L. Skin as a living coloring book: how epithelial cells create patterns of pigmentation. *Pigment Cell Melanoma Res.* **27,** 1014-31 (2014).
20. Szabó, G., Gerald, A. B., Pathak, M. A. & Fitzpatrick, T. B. Racial differences in the fate of melanosomes in human epidermis. *Nature* **222,** 1081–2 (1969).
21. Rees, J. L. Genetics of hair and skin color. *Annu. Rev. Genet.* **37,** 67–90 (2003).
22. Alaluf, S. *et al.* The impact of epidermal melanin on objective measurements of human skin colour. *Pigment Cell Res.* **15,** 119–26 (2002).
23. Sturm, R. A. & Frudakis, T. N. Eye colour: portals into pigmentation genes and ancestry. **20,** 327-32 (2004).
24. Swan, G. A. Chemical structure of melanins. *Ann. N. Y. Acad. Sci.* **100,** 1005–19 (1963).
25. Nicolaus, R. A., Piattelli, M. & Fattorusso, E. The structure of melanins and melanogenesis. IV. On some natural melanins. *Tetrahedron* **20,** 1163–72 (1964).
26. Raper, H. S. The Tyrosinase-tyrosine Reaction: Production from Tyrosine of 5: 6-Dihydroxyindole and 5: 6-Dihydroxyindole-2-carboxylic Acid-the Precursors of Melanin. *Biochem. J.* **21,** 89–96 (1927).
27. Mason, H. S. THE CHEMISTRY OF MELANIN. III. MECHANISM OF THE OXIDATION OF DIHYDROXYPHENYLALANINE BY TYROSINASE. *J. Biol. Chem.* **172,** 83–99 (1948).
28. Ito, S. & Wakamatsu, K. Chemistry of mixed melanogenesis - Pivotal roles of dopaquinone.

Photochemistry and Photobiology **84,** 582–592 (2008).
29. Lamoreux, M. L., Wakamatsu, K. & Ito, S. Interaction of Major Coat Color Gene Functions in Mice as Studied by Chemical Analysis of Eumelanin and Pheomelanin. *Pigment Cell Res.* **14,** 23–31 (2001).
30. Ito, S. The IFPCS presidential lecture: a chemist's view of melanogenesis. *Pigment Cell Res.* **16,** 230–6 (2003).
31. Wakamatsu, K., Ohtara, K. & Ito, S. Chemical analysis of late stages of pheomelanogenesis: conversion of dihydrobenzothiazine to a benzothiazole structure. *Pigment Cell Melanoma Res.* **22,** 474–86 (2009).
32. Land, E. J. & Riley, P. a. Spontaneous redox reactions of dopaquinone and the balance between the eumelanic and phaeomelanic pathways. *Pigment Cell Res.* **13,** 273–7 (2000).
33. Land, E. J., Ito, S., Wakamatsu, K. & Riley, P. A. Rate Constants for the First Two Chemical Steps of Eumelanogenesis. *Pigment Cell Res.* **16,** 487–493 (2003).
34. Land, E. J., Ramsden, C. a & Riley, P. a. Pulse radiolysis studies of ortho-quinone chemistry relevant to melanogenesis. *J. Photochem. Photobiol. B.* **64,** 123–35 (2001).
35. Jara, J. R., Aroca, P., Solano, F., Martinez, J. H. & Lozano, J. A. The role of sulfhydryl compounds in mammalian melanogenesis: the effect of cysteine and glutathione upon tyrosinase and the intermediates of the pathway. *Biochim. Biophys. Acta* **967,** 296–303 (1988).
36. Fitzpatrick, T. B., Becker, S. W., Lerner, A. B. & Montgomery, H. Tyrosinase in Human Skin: Demonstration of Its Presence and of Its Role in Human Melanin Formation. *Science* **112,** 223–225 (1950).
37. Del Marmol, V. *et al.* Glutathione Depletion Increases Tyrosinase Activity in Human Melanoma Cells. *J. Invest. Dermatol.* **101,** 871–874 (1993).
38. Tsukamoto, K., Jackson, I. J., Urabe, K., Montague, P. M. & Hearing, V. J. A second tyrosinase-related protein, TRP-2, is a melanogenic enzyme termed DOPAchrome tautomerase. *EMBO J.* **11,** 519–26 (1992).
39. Orlow, S. J. *et al.* High-Molecular-Weight Forms of Tyrosinase and the Tyrosinase-Related Proteins: Evidence for a Melanogenic Complex. *J. Invest. Dermatol.* **103,** 196–201 (1994).
40. Alberts, B. *Molecular biology of the cell*. (Garland Science, 2002).
41. Watson, J. D. & Crick, F. H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18,** 123–31 (1953).
42. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122,** 565–81 (2008).
43. Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *J. Exp. Med.* **79,** 137–58 (1944).
44. Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10,** 654–9 (1951).
45. Hershey, A. D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* **36,** 39–56 (1952).
46. Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184,** 868–71 (1974).
47. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421,** 448–53 (2003).
48. Kornberg, R. D. & Lorch, Y. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. *Cell* **98,** 285–294 (1999).
49. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152,** 1237–51 (2013).
50. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157,** 77–94 (2014).
51. Haraksingh, R. R. & Snyder, M. P. Impacts of variation in the human genome on gene regulation. *J. Mol. Biol.* **425,** 3970–7 (2013).
52. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
53. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42,** D1001–6 (2014).
54. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark

road from association to function. *Am. J. Hum. Genet.* **93,** 779–97 (2013).
55. Hao, K., Schadt, E. E. & Storey, J. D. Calibrating the performance of SNP arrays for whole-genome association studies. *PLoS Genet.* **4,** e1000109 (2008).
56. Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat. Rev. Genet.* **12,** 628–40 (2011).
57. Visel, A., Rubin, E. M. & Pennacchio, L. a. Genomic views of distant-acting enhancers. *Nature* **461,** 199–205 (2009).
58. Smith, A. G. *et al.* The human melanocortin-1 receptor locus: analysis of transcription unit, locus polymorphism and haplotype evolution. *Gene* **281,** 81–94 (2001).
59. Rana, B. K. *et al.* High polymorphism at the human melanocortin 1 receptor locus. *Genetics* **151,** 1547–57 (1999).
60. Sturm, R. A. *et al.* Genetic association and cellular function of MC1R variant alleles in human pigmentation. *Ann. N. Y. Acad. Sci.* **994,** 348–58 (2003).
61. Searle, A. G. *Comparative genetics of coat colour in mammals*. (Logos Press, London; Academic Press, New York, 1968).
62. Robbins, L. S. *et al.* Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* **72,** 827–34 (1993).
63. Valverde, P., Healy, E., Jackson, I., Rees, J. L. & Thody, A. J. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat. Genet.* **11,** 328–30 (1995).
64. Beaumont, K. a, Shekar, S. N., Cook, A. L., Duffy, D. L. & Sturm, R. a. Red hair is the null phenotype of MC1R. *Hum. Mutat.* **29,** E88–94 (2008).
65. Lin, J. Y. & Fisher, D. E. Melanocyte biology and skin pigmentation. *Nature* **445,** 843–50 (2007).
66. Flanagan, N. *et al.* Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. *Hum. Mol. Genet.* **9,** 2531–7 (2000).
67. Harding, R. M. *et al.* Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66,** 1351–61 (2000).
68. Makova, K. & Norton, H. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. *Peptides* **26,** 1901–8 (2005).
69. Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310,** 1782–6 (2005).
70. Cook, A. L. *et al.* Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. *J. Invest. Dermatol.* **129,** 392–405 (2009).
71. Sturm, R. a. Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* **18,** R9–17 (2009).
72. Grønskov, K., Ek, J. & Brondum-Nielsen, K. Oculocutaneous albinism. *Orphanet J. Rare Dis.* **2,** 43 (2007).
73. Grønskov, K. *et al.* Mutations in c10orf11, a melanocyte-differentiation gene, cause autosomal-recessive albinism. *Am. J. Hum. Genet.* **92,** 415–21 (2013).
74. Kausar, T., Bhatti, M. A., Ali, M., Shaikh, R. S. & Ahmed, Z. M. OCA5, a novel locus for non-syndromic oculocutaneous albinism, maps to chromosome 4q24. *Clin. Genet.* **84,** 91–3 (2013).
75. Mártinez-García, M. & Montoliu, L. Albinism in Europe. *J. Dermatol.* **40,** 319–24 (2013).
76. Montoliu, L. *et al.* Increasing the complexity: new genes and new types of albinism. *Pigment Cell Melanoma Res.* **27,** 11–8 (2014).
77. Inagaki, K. *et al.* Oculocutaneous albinism type 4 is one of the most common types of albinism in Japan. *Am. J. Hum. Genet.* **74,** 466–71 (2004).
78. Rundshagen, U., Zühlke, C., Opitz, S., Schwinger, E. & Käsmann-Kellner, B. Mutations in the MATP gene in five German patients affected by oculocutaneous albinism type 4. *Hum. Mutat.* **23,** 106–10 (2004).
79. Newton, J. M. *et al.* Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4. *Am. J. Hum. Genet.* **69,** 981–8 (2001).
80. Soejima, M., Tachida, H., Ishida, T., Sano, A. & Koda, Y. Evidence for recent positive selection at the human AIM1 locus in a European population. *Mol. Biol. Evol.* **23,** 179–88 (2006).
81. Cullinane, A. R. *et al.* Homozygosity mapping and whole-exome sequencing to detect SLC45A2 and G6PC3 mutations in a single patient with oculocutaneous albinism and neutropenia. *J. Invest. Dermatol.* **131,** 2017–25 (2011).
82. Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color

and skin pigmentation. *PLoS Genet.* **4,** e1000074 (2008).
83. Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39,** 1443–52 (2007).
84. Liu, F. *et al.* Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet.* **6,** e1000934 (2010).
85. Duffy, D. L. *et al.* Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. *J. Invest. Dermatol.* **130,** 520–8 (2010).
86. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6,** e1000993 (2010).
87. Sturm, R. A. Human "coat colour" genetics. *Pigment Cell Melanoma Res.* **21,** 115–6 (2008).
88. King, R. A. *et al.* Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype. *Hum. Genet.* **113,** 502–13 (2003).
89. Kamaraj, B. & Purohit, R. Mutational analysis of oculocutaneous albinism: a compact review. *Biomed Res. Int.* **2014,** 905472 (2014).
90. Sulem, P. *et al.* Two newly identified genetic determinants of pigmentation in Europeans. *Nat. Genet.* **40,** 835–7 (2008).
91. Branicki, W. *et al.* Model-based prediction of human hair color using DNA variants. *Hum. Genet.* **129,** 443–54 (2011).
92. Pośpiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A. & Branicki, W. Gene-gene interactions contribute to eye colour variation in humans. *J. Hum. Genet.* **56,** 447–55 (2011).
93. Shriver, M. D. *et al.* Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum. Genet.* **112,** 387–99 (2003).
94. Stokowski, R. P. *et al.* A genomewide association study of skin pigmentation in a South Asian population. *Am. J. Hum. Genet.* **81,** 1119–32 (2007).
95. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat. Genet.* **44,** 676–80 (2012).
96. Jacobs, L. C. *et al.* Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Hum. Genet.* **132,** 147–58 (2013).
97. Morice-Picard, F. *et al.* 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 Melanoma Res.* **27,** 59–71 (2014).
98. Simeonov, D. R. *et al.* DNA variations in oculocutaneous albinism: an updated mutation list and current outstanding issues in molecular diagnostics. *Hum. Mutat.* **34,** 827–35 (2013).
99. Buiting, K. Prader-Willi syndrome and Angelman syndrome. *Am. J. Med. Genet. C. Semin. Med. Genet.* **154C,** 365–76 (2010).
100. Hirobe, T. How are proliferation and differentiation of melanocytes regulated? *Pigment Cell Melanoma Res.* **24,** 462–78 (2011).
101. Rosemblat, S. *et al.* Identification of a melanosomal membrane protein encoded by the pink-eyed dilution (type II oculocutaneous albinism) gene. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12071–75 (1994).
102. Puri, N., Gardner, J. M. & Brilliant, M. H. Aberrant pH of melanosomes in pink-eyed dilution (p) mutant melanocytes. *J. Invest. Dermatol.* **115,** 607–13 (2000).
103. Manga, P., Boissy, R. E., Pifko-Hirst, S., Zhou, B. K. & Orlow, S. J. Mislocalization of melanosomal proteins in melanocytes from mice with oculocutaneous albinism type 2. *Exp. Eye Res.* **72,** 695–710 (2001).
104. Potterf, S. B. *et al.* Normal tyrosine transport and abnormal tyrosinase routing in pink-eyed dilution melanocytes. *Exp. Cell Res.* **244,** 319–26 (1998).
105. Sitaram, A. *et al.* Localization to mature melanosomes by virtue of cytoplasmic dileucine motifs is required for human OCA2 function. *Mol. Biol. Cell* **20,** 1464–77 (2009).
106. Sitaram, A. *et al.* Differential recognition of a dileucine-based sorting signal by AP-1 and AP-3 reveals a requirement for both BLOC-1 and AP-3 in delivery of OCA2 to melanosomes. *Mol. Biol. Cell* **23,** 3178–92 (2012).
107. Staleva, L., Manga, P. & Orlow, S. J. Pink-eyed Dilution Protein Modulates Arsenic Sensitivity and Intracellular Glutathione Metabolism. **13,** 4206–4220 (2002).
108. Cheng, T., Orlow, S. J. & Manga, P. Loss of Oca2 disrupts the unfolded protein response and increases resistance to endoplasmic reticulum stress in melanocytes. *Pigment Cell Melanoma*

Res. **26,** 826–34 (2013).
109. Chen, K., Manga, P. & Orlow, S. J. Pink-eyed dilution protein controls the processing of tyrosinase. *Mol. Biol. Cell* **13,** 1953–64 (2002).
110. Branicki, W., Brudnik, U. & Wojas-Pelc, A. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. *Ann. Hum. Genet.* **73,** 160–70 (2009).
111. Eiberg, H. *et al.* 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. Genet.* **123,** 177–87 (2008).
112. Frudakis, T., Terravainen, T. & Thomas, M. Multilocus OCA2 genotypes specify human iris colors. *Hum. Genet.* **122,** 311–26 (2007).
113. Kayser, M. *et al.* Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am. J. Hum. Genet.* **82,** 411–23 (2008).
114. Lao, O., de Gruijter, J. M., van Duijn, K., Navarro, A. & Kayser, M. Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann. Hum. Genet.* **71,** 354–69 (2007).
115. Nan, H. *et al.* Genome-wide association study of tanning phenotype in a population of European ancestry. *J. Invest. Dermatol.* **129,** 2250–7 (2009).
116. Shekar, S. N. *et al.* Linkage and association analysis of spectrophotometrically quantified hair color in Australian adolescents: the effect of OCA2 and HERC2. *J. Invest. Dermatol.* **128,** 2807–14 (2008).
117. Sturm, R. a *et al.* A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82,** 424–31 (2008).
118. Liu, F. *et al.* Eye color and the prediction of complex phenotypes from genotypes. *Curr. Biol.* **19,** R192–3 (2009).
119. Walsh, S. *et al.* Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. *Forensic Sci. Int. Genet.* **5,** 464–71 (2011).
120. Hider, J. L. *et al.* Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol. Biol.* **13,** 150 (2013).
121. Smyth, I. M. *et al.* Genomic anatomy of the Tyrp1 (brown) deletion complex. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 3704–9 (2006).
122. Vanhoutteghem, A. *et al.* Basonuclin 2 has a function in the multiplication of embryonic craniofacial mesenchymal cells and is orthologous to disco proteins. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 14432–7 (2009).
123. Lang, M. R., Patterson, L. B., Gordon, T. N., Johnson, S. L. & Parichy, D. M. Basonuclin-2 requirements for zebrafish adult pigment pattern development and female fertility. *PLoS Genet.* **5,** e1000744 (2009).
124. Patterson, L. B. & Parichy, D. M. Interactions with iridophores and the tissue environment required for patterning melanophores and xanthophores during zebrafish adult pigment stripe formation. *PLoS Genet.* **9,** e1003561 (2013).
125. Vanhoutteghem, A. & Djian, P. Basonuclin 2: an extremely conserved homolog of the zinc finger protein basonuclin. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3468–73 (2004).
126. Hervé, F., Vanhoutteghem, A. & Djian, P. [Basonuclins and DISCO proteins: regulators of development in vertebrates and insects]. *Med. Sci. (Paris).* **28,** 55–61 (2012).
127. Paun, A. & Pitha, P. M. The IRF family, revisited. *Biochimie* **89,** 744–53 (2007).
128. Gualco, G., Weiss, L. M. & Bacchi, C. E. MUM1/IRF4: A Review. *Appl. Immunohistochem. Mol. Morphol.* **18,** 301–10 (2010).
129. Sundram, U., Harvell, J. D., Rouse, R. V & Natkunam, Y. Expression of the B-cell proliferation marker MUM1 by melanocytic lesions and comparison with S100, gp100 (HMB45), and MelanA. *Mod. Pathol.* **16,** 802–10 (2003).
130. Cario-André, M., Pain, C., Gauthier, Y., Casoli, V. & Taieb, A. In vivo and in vitro evidence of dermal fibroblasts influence on human epidermal pigmentation. *Pigment Cell Res.* **19,** 434–42 (2006).
131. Trantow, C. M., Cuffy, T. L., Fingert, J. H., Kuehn, M. H. & Anderson, M. G. Microarray analysis of iris gene expression in mice with mutations influencing pigmentation. *Invest. Ophthalmol. Vis. Sci.* **52,** 237–48 (2011).
132. Gutiérrez-Gil, B., Wiener, P. & Williams, J. L. Genetic effects on coat colour in cattle: dilution of eumelanin and phaeomelanin pigments in an F2-Backcross Charolais x Holstein population.

BMC Genet. **8,** 56 (2007).
133. Furumura, M., Sakai, C., Abdel-Malek, Z., Barsh, G. S. & Hearing, V. J. The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. *Pigment Cell Res.* **9,** 191–203 (1996).
134. Suzuki, I. *et al.* Agouti signaling protein inhibits melanogenesis and the response of human melanocytes to alpha-melanotropin. *J. Invest. Dermatol.* **108,** 838–842 (1997).
135. Kanetsky, P. A. *et al.* A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *Am. J. Hum. Genet.* **70,** 770–5 (2002).
136. Bonilla, C. *et al.* The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. *Hum. Genet.* **116,** 402–6 (2005).
137. Gudbjartsson, D. F. *et al.* ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat. Genet.* **40,** 886–91 (2008).
138. Maccioni, L. *et al.* Variants at chromosome 20 (ASIP locus) and melanoma risk. *Int. J. Cancer* **132,** 42–54 (2013).
139. Brown, K. M. *et al.* Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat. Genet.* **40,** 838–40 (2008).
140. Liu, F., Wen, B. & Kayser, M. Colorful DNA polymorphisms in humans. *Semin. Cell Dev. Biol.* **24,** 562–75 (2013).
141. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq : a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10,** 57–63 (2009).
142. Ramsköld, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **5,** e1000598 (2009).
143. Zubakov, D., Hanekamp, E. & Kayser, M. Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. *Int. J. Legal Med.* **122,** 135–142 (2008).
144. Haas, C., Klesser, B., Maake, C., Bär, W. & Kratzer, a. mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Sci. Int. Genet.* **3,** 80–8 (2009).
145. Juusola, J. & Ballantyne, J. Multiplex mRNA profiling for the identification of body fluids. *Forensic Sci. Int.* **152,** 1–12 (2005).
146. Visser, M., Zubakov, D., Ballantyne, K. N. & Kayser, M. MRNA-based skin identification for forensic applications. *Int. J. Legal Med.* **125,** 253–263 (2011).
147. Schallreuter, K. U., Kothari, S., Chavan, B. & Spencer, J. D. Regulation of melanogenesis--controversies and new concepts. *Exp. Dermatol.* **17,** 395–404 (2008).
148. Levy, C., Khaled, M. & Fisher, D. E. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol. Med.* **12,** 406–14 (2006).
149. Law, M. H., Macgregor, S. & Hayward, N. K. Melanoma genetics: recent findings take us beyond well-traveled pathways. *J. Invest. Dermatol.* **132,** 1763–74 (2012).
150. Strub, T. *et al.* Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30,** 2319–32 (2011).
151. Hoek, K. S. MITF: the power and the glory. *Pigment Cell Melanoma Res.* **24,** 262–3 (2011).
152. Mountjoy, K. G., Robbins, L. S., Mortrud, M. T. & Cone, R. D. The cloning of a family of genes that encode the melanocortin receptors. *Science* **257,** 1248–51 (1992).
153. D'Orazio, J. a *et al.* Topical drug rescue strategy and skin protection based on the role of Mc1r in UV-induced tanning. *Nature* **443,** 340–4 (2006).
154. Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 12955–61 (2007).
155. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11,** 761–72 (2010).
156. Thomas, M. C. & Chiang, C.-M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41,** 105–78 (2006).
157. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36,** 541–6 (2009).
158. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13,** 720–31 (2012).
159. Price, D. H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* **20,** 2629–34 (2000).
160. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental

control. *Nat. Rev. Genet.* **13,** 613–26 (2012).
161. Lelli, K. M., Slattery, M. & Mann, R. S. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu. Rev. Genet.* **46,** 43–68 (2012).
162. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10,** 252–63 (2009).
163. Villar, D., Flicek, P. & Odom, D. T. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat. Rev. Genet.* **15,** 221–33 (2014).
164. Maston, G. A., Landt, S. G., Snyder, M. & Green, M. R. Characterization of enhancer function from genome-wide analyses. *Annu. Rev. Genomics Hum. Genet.* **13,** 29–57 (2012).
165. Hsiao, J. J. & Fisher, D. E. The roles of microphthalmia-associated transcription factor and pigmentation in melanoma. *Arch. Biochem. Biophys.* **563,** 28-34 (2014).
166. Becker, P. B. & Hörz, W. ATP-dependent nucleosome remodeling. *Annu. Rev. Biochem.* **71,** 247–73 (2002).
167. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15,** 272–86 (2014).
168. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–5 (2000).
169. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120,** 169–81 (2005).
170. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039–43 (2002).
171. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18,** 1251–62 (2004).
172. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–6 (2010).
173. Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12,** 283–93 (2011).
174. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308 (1981).
175. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
176. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
177. Marsman, J. & Horsfield, J. A. Long distance relationships: enhancer-promoter communication and dynamic gene transcription. *Biochim. Biophys. Acta* **1819,** 1217–27 (2012).
178. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13,** 59–69 (2012).
179. Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57,** 159–97 (1988).
180. Bartkuhn, M. & Renkawitz, R. Long range chromatin interactions involved in gene regulation. *Biochim. Biophys. Acta* **1783,** 2161–6 (2008).
181. Palstra, R.-J. T. S. Close encounters of the 3C kind: long-range chromatin interactions and transcriptional regulation. *Brief. Funct. Genomic. Proteomic.* **8,** 297–309 (2009).
182. Merkenschlager, M. & Odom, D. T. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell* **152,** 1285–97 (2013).
183. Palstra, R.-J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35,** 190–4 (2003).
184. Cook, P. R. A model for all genomes: the role of transcription factories. *J. Mol. Biol.* **395,** 1–10 (2010).
185. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43,** 630–8 (2011).
186. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20,** 2349–54 (2006).
187. Arnone, M. I. & Davidson, E. H. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124,** 1851–64 (1997).
188. Murisier, F., Guichard, S. & Beermann, F. The tyrosinase enhancer is activated by Sox10 and Mitf in mouse melanocytes. *Pigment Cell Res.* **20,** 173–84 (2007).

189. Hardison, R. C. & Taylor, J. Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.* **13,** 469–83 (2012).
190. Kellis, M. *et al.* Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6131–8 (2014).
191. Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13,** 840–52 (2012).
192. Fisher, W. W. *et al.* DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21330–5 (2012).
193. Ludwig, M. Z. *et al.* Functional evolution of a cis-regulatory module. *PLoS Biol.* **3,** e93 (2005).
194. De Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26,** 11–24 (2012).
195. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–11 (2002).
196. Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2,** 1722–33 (2007).
197. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9,** e1001046 (2011).
198. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28,** 1045–8 (2010).
199. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507,** 455–61 (2014).
200. Gorkin, D. U. *et al.* Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Res.* **22,** 2290–301 (2012).
201. Kleinjan, D. A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76,** 8–32 (2005).
202. Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9742–6 (2010).
203. Jia, L. *et al.* Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet.* **5,** e1000597 (2009).
204. Pomerantz, M. M. *et al.* The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.* **41,** 882–4 (2009).
205. Harismendy, O. *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* **470,** 264–8 (2011).
206. Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21,** 3255–63 (2012).
207. Tuupanen, S. *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.* **41,** 885–90 (2009).
208. Van den Boogaard, M. *et al.* Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J. Clin. Invest.* **122,** 2519–30 (2012).
209. Wasserman, N. F., Aneas, I. & Nobrega, M. A. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res.* **20,** 1191–7 (2010).
210. Wright, J. B., Brown, S. J. & Cole, M. D. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol. Cell. Biol.* **30,** 1411–20 (2010).
211. Birney, E., Lieb, J. D., Furey, T. S., Crawford, G. E. & Iyer, V. R. Allele-specific and heritable chromatin signatures in humans. *Hum. Mol. Genet.* **19,** R204–9 (2010).
212. McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328,** 235–9 (2010).
213. Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A. & Kingsley, D. M. A molecular basis for classic blond hair color in Europeans. *Nat. Genet.* **46,** 748–52 (2014).
214. Miller, C. T. *et al.* cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131,** 1179–89 (2007).
215. Wehrle-haller, B. The Role of Kit-Ligand in Melanocyte Development and Epidermal Homeostasis. *Pigment Cell Res.* **16,** 287–296 (2003).
216. Picardo, M. & Cardinali, G. The genetic determination of skin pigmentation: KITLG and the KITLG/c-Kit pathway as key players in the onset of human familial pigmentary diseases. *J. Invest.*

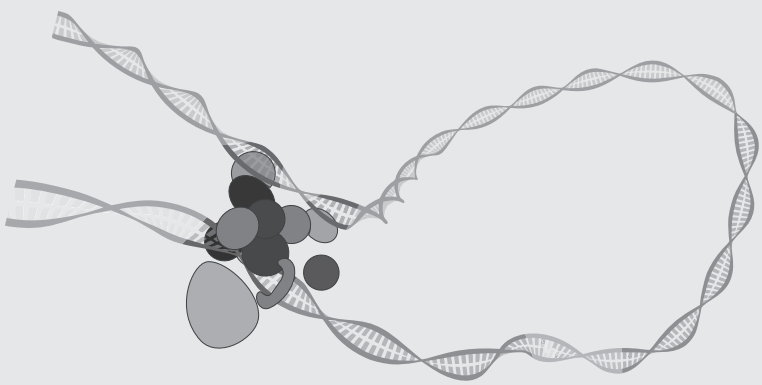
Dermatol. **131,** 1182–5 (2011).
217. Amyere, M. *et al.* KITLG mutations cause familial progressive hyper- and hypopigmentation. *J. Invest. Dermatol.* **131,** 1234–9 (2011).
218. Do, T. N., Ucisik-Akkaya, E., Davis, C. F., Morrison, B. A. & Dorak, M. T. An intronic polymorphism of IRF4 gene influences gene transcription in vitro and shows a risk association with childhood acute lymphoblastic leukemia in males. *Biochim. Biophys. Acta* **1802,** 292–300 (2010).
219. Praetorius, C. *et al.* A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155,** 1022–33 (2013).

Chapter 2

HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the _OCA2_ promoter

Mijke Visser, Manfred Kayser, and Robert-Jan Palstra

Genome Research, 3, 446-455 (2012)



Abstract

Pigmentation of skin, eye and hair reflects some of the most evident common phenotypes in humans. Several candidate genes for human pigmentation are identified. The SNP rs12913832 has strong statistical association with human pigmentation. It is located within an intron of the non-pigment gene *HERC2*, 21 kb upstream of the pigment gene *OCA2*, and the region surrounding rs12913832 is highly conserved among animal species. However, the exact functional role of *HERC2* rs12913832 in human pigmentation is unknown. Here we demonstrate that the *HERC2* rs12913832 region functions as an enhancer regulating *OCA2* transcription. In darkly pigmented human melanocytes carrying the rs12913832 T-allele, we detected binding of the transcription factors HLTF, LEF1 and MITF to the *HERC2* rs12913832 enhancer, and a long-range chromatin loop between this enhancer and the *OCA2* promoter which leads to elevated *OCA2* expression. In contrast, in lightly pigmented melanocytes carrying the rs12913832 C-allele, chromatin-loop formation, transcription factor recruitment, and *OCA2* expression are all reduced. Hence, we demonstrate that allelic variation of a common noncoding SNP located in a distal regulatory element not only disrupts the regulatory potential of this element but also affects its interaction with the relevant promoter. We provide the key mechanistic insight that allele-dependent differences in chromatin-loop formation (i.e., structural differences in the folding of gene loci) results in differences in allelic gene expression that affects common phenotypic traits. This concept is highly relevant for future studies aiming to unveil the functional basis of genetically-determined phenotypes, including diseases.

Introduction

Genome-wide association studies (GWAS) have successfully identified numerous genetic variants that are significantly associated with complex diseases or other phenotypic traits (Manolio 2010). Intriguingly, many of the identified disease or other phenotype-associated DNA variants are located in non-coding regions of the human genome. It remains largely unknown if and how these noncoding DNA variants influence phenotypic traits at the molecular level. Although linkage with unknown causal DNA variants is often assumed as a major factor to explain association of noncoding DNA variants, it has been suggested that some of these noncoding DNA variants constitute potential regulatory elements for distal genes (Coetzee et al. 2010; Freedman et al.). However, so far, this has only been demonstrated for some low-penetrance disease susceptibility loci (Ahmadiyeh et al. 2010; Harismendy et al. 2010; Jia et al. 2009; Pomerantz et al. 2009; Wright et al. 2010).

Pigmentation of skin, eye and hair reflects one of the clearest common phenotypes in humans, but is in fact a complex polygenetic trait. Several genes are involved in the human pigmentation pathway and many DNA polymorphisms are associated with human pigmentation variation (Sturm 2009). The SNP rs12913832 is highly associated with human eye (Eiberg et al. 2008; Liu et al. 2009; Liu et al. 2010; Sturm et al. 2008), hair (Branicki et al. 2009; Branicki et al. 2011; Han et al. 2008), and skin (Branicki et al. 2009; Spichenok et al. 2010; Valenzuela et al. 2010) color variation. This SNP is the key determinant of human eye color variation and has strong predictive power for human pigmentation (Branicki et

al. 2011; Liu et al. 2009; Liu et al. 2010; Spichenok et al. 2010; Valenzuela et al. 2010). Interestingly, rs12913832 is not located in a pigment gene, but in intron 86 of the *HERC2* gene, 21 kb upstream of the promoter of *OCA2* (Fig. 1A). The *OCA2* gene is involved in human pigmentation since it regulates melanosomal pH, and mutations in *OCA2* cause oculocutaneous albinism type II (Brilliant 2001). The region directly surrounding *HERC2* rs12913832 is highly conserved among animal species (Fig. 1A; Supplemental Table S1; Sturm et al. 2008), and this SNP strongly associates with *OCA2* expression levels (Cook et al. 2009). However, if and how *HERC2* rs12913832 directly regulates human pigmentation is currently unknown. It was previously hypothesized that this SNP acts as a distal regulatory region that either silences or enhances (Eiberg et al. 2008; Sturm et al. 2008; Sturm and Larsson 2009) *OCA2* expression, but convincing experimental data that would confirm or reject either one of these hypotheses remains to be presented.

In this study, we used various molecular approaches to investigate the functional role of *HERC2* rs12913832 in human pigmentation variation. In darkly pigmented and lightly pigmented melanocyte cell lines with different genotypes for *HERC2* rs12913832, we investigated the relation between the *HERC2* rs12913832 genotype and *OCA2* expression, surveyed the epigenomic landscape of the region around *HERC2* rs12913832, and mapped the chromatin folding of the *OCA2*-*HERC2* locus. Our study provides crucial mechanistic insight into how a noncoding DNA polymorphism can affect a common human phenotypic trait by modulating the chromatin folding of a gene locus.

Results

Characterization and suitability of the melanocyte cell system

To investigate the potential functional role of *HERC2* rs12913832 we used commercially available primary Human Epidermal Melanocytes of neonatal origin derived from a lightly pigmented (HEMn-LP) and a darkly pigmented donor (HEMn-DP) (Cascade Biologics, Invitrogen). The MCF7 breast-cancer cell line was used for control purposes. Genotyping, either by iPLEX (Sequenom) (Liu et al. 2009) or direct DNA sequencing of PCR products, of SNPs statistically associated with human pigmentation revealed that HEMn-LP cells are homozygous for the rs12913832 C-allele, while HEMn-DP cells are homozygous for the rs12913832 T-allele (Fig. 1B). STRUCTURE analysis (Pritchard et al. 2000) of DNA samples from HEMn-LP and HEMn-DP cells using 24 ancestry-informative autosomal SNPs together with a worldwide reference dataset (HGDP-CEPH) (Lao et al. 2010), revealed a probable European and a probable Sub-Saharan African genetic origin of the respective cell donors (Supplemental Fig. S1). Both datasets are in agreement with the light and dark skin-color phenotype information of the HEMn-LP and HEMn-DP cell donors, as provided by Cascade Biologics.

We investigated the transcription status of the *OCA2* gene in our skin melanocyte system by measuring primary *OCA2*-transcripts using quantitative (qRT-PCR). We observed higher expression of *OCA2* in HEMn-DP cells, carrying the *HERC2* rs12913832 T-allele, than in HEMn-LP cells carrying the C-allele (Fig. 1C; $p < 0.01$), confirming previous observations

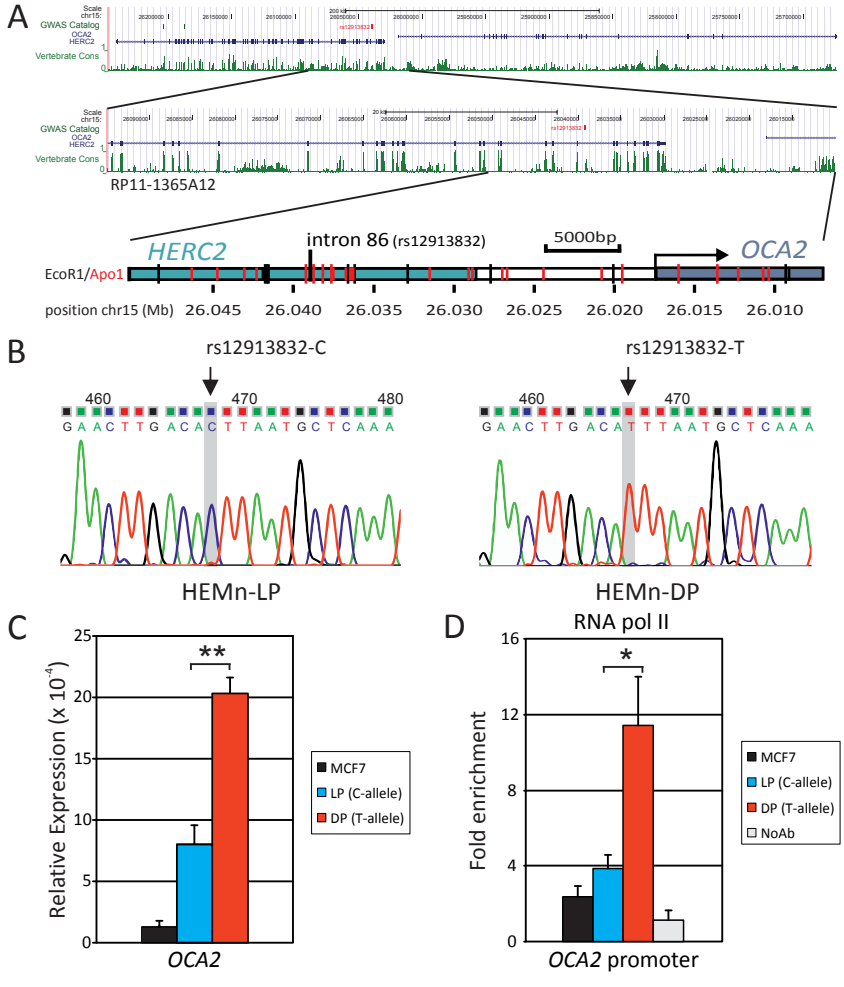


Figure 1. Characterization and suitability of the HEMn cell system: the *OCA2* gene is differentially expressed in HEMn-LP and HEMn-DP cells. (A) UCSC browser NCBI36/hg18 assembly (http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg18) overview of the *OCA2-HERC2* locus (*top* panel). (*Middle* panel) The region covered on BAC RP11-1365A12. Vertebrate conservation (green); the position of rs12913832 (red). (*Lower* panel) A schematic overview of the region investigated in this study. Restriction enzyme digestion sites are indicated. (B) Sequence analysis of the region around *HERC2* rs12913832 in HEMn-LP (*left*) and HEMn-DP (*right*). The genotypes of rs12913832 are determined by direct sequencing of PCR fragments containing rs12913832. (C) RT-qPCR analysis of *OCA2* primary transcripts in MCF7 and HEMn cells demonstrates differential *OCA2* expression between HEMn-LP and HEMn-DP cells. Each gene-expression analysis is carried out in triplicate and normalized to an endogenous reference gene (*ACTB*). (D) ChIP-qPCR of RNA pol II binding at the *OCA2* promoter in MCF7, HEMn-LP and HEMn-DP cells. Enrichment is calculated relative to necdin (*NDN*), and values are normalized to input measurements. All ChIP analyses are performed in triplicate. Data are represented as mean ± SEM; (*) $p < 0.05$; (**) $p < 0.01$.

(Cook et al. 2009). We also tested for the presence of RNA polymerase II on the *OCA2* promoter using chromatin immuno precipitation (ChIP) using an antibody against the POLR2A subunit of RNA polymerase II followed by qPCR (ChIP-qPCR). ChIP is a technology that identifies chromatin regions bound by a specific protein (i.e., a transcription factor) using an antibody specific to the protein of interest to enrich cross-linked fragmented chromatin bound by such a protein from the bulk of the chromatin. We observed a higher RNA pol II recruitment to the *OCA2* promoter in HEMn-DP cells than in HEMn-LP cells (Fig. 1D; $p < 0.05$). Both *OCA2* gene expression and RNA polymerase II recruitment were very low in MCF7 control cells (Fig. 1C,D). From these results, we conclude that the HEMn-LP/HEMn-DP melanocyte system constitutes an appropriate system to investigate the functional effect of *HERC2* rs12913832 on human pigmentation.

Identification of HERC2 rs12913832 as a melanocyte-specific enhancer

HERC2 rs12913832 was previously hypothesized to function as a regulatory region (Sturm et al. 2008;Cook et al. 2009). Gene regulatory regions are characterized by open chromatin and active chromatin marks (Giresi et al. 2007). Hence we used formaldehyde-assisted identification of regulatory elements (FAIRE) and ChIP-qPCR of histone modifications, to investigate the chromatin status of the *HERC2* rs12913832 region. FAIRE identifies regulatory DNA regions based on differences in cross-link ability between chromatin of regulatory DNA regions and the bulk of the chromatin (Giresi et al. 2007). FAIRE analysis revealed an open chromatin structure in the region around *HERC2* rs12913832 in both HEMn-LP and HEMn-DP cells, while this region was shown to be closed in MCF7 control cells (Fig. 2A). In addition, the level of enrichment identified by this assay was higher for HEMn-DP cells than for HEMn-LP cells (Fig. 2A; $p < 0.05$), suggesting that the chromatin of the *HERC2* rs12913832 region is more open when the rs12913832 T-allele is present relative to when the C-allele is present. ChIP-qPCR analysis detected elevated levels of acetylated histone H3, an active chromatin mark, at *HERC2* rs12913832 in HEMn-LP and HEMn-DP cells, but not in the MCF7 control cells (Fig. 2B). Both observations are consistent with the hypothesis that *HERC2* rs12913832 acts as a human melanocyte-specific gene regulatory region.

Enhancers are distal gene regulatory elements that act positively on gene expression and are enriched for mono-methylation of lysine 4 on histone H3 (Heintzman et al. 2007). Recently it was shown that active enhancers additionally accumulate acetylation of lysine 27 on histone H3 (Creyghton et al. 2010; Rada-Iglesias et al. 2010). ChIP-qPCR analysis of these chromatin marks revealed the presence of both enhancer marks on rs12913832 in both HEMn cell types, but not in MCF7 control cells (Fig. 2C,D). Similar to the differences observed with the FAIRE assay, higher enrichments for the chromatin marks were found on *HERC2* rs12913832 in HEMn-DP cells than in HEMn-LP cells, although these differences do not reach statistical significance (Fig. 2B-D). The above observations are consistent with the *HERC2* rs12913832 region operating as an enhancer in skin melanocytes. Interestingly, the level of the enhancer marks in the *HERC2* rs12913832 alleles also positively correlates with *OCA2* expression and the level of melanocyte pigmentation.

We further investigated the enhancer potential of the region encompassing *HERC2* rs12913832 by cloning this region from HEMn-LP and HEMn-DP cells into a luciferase

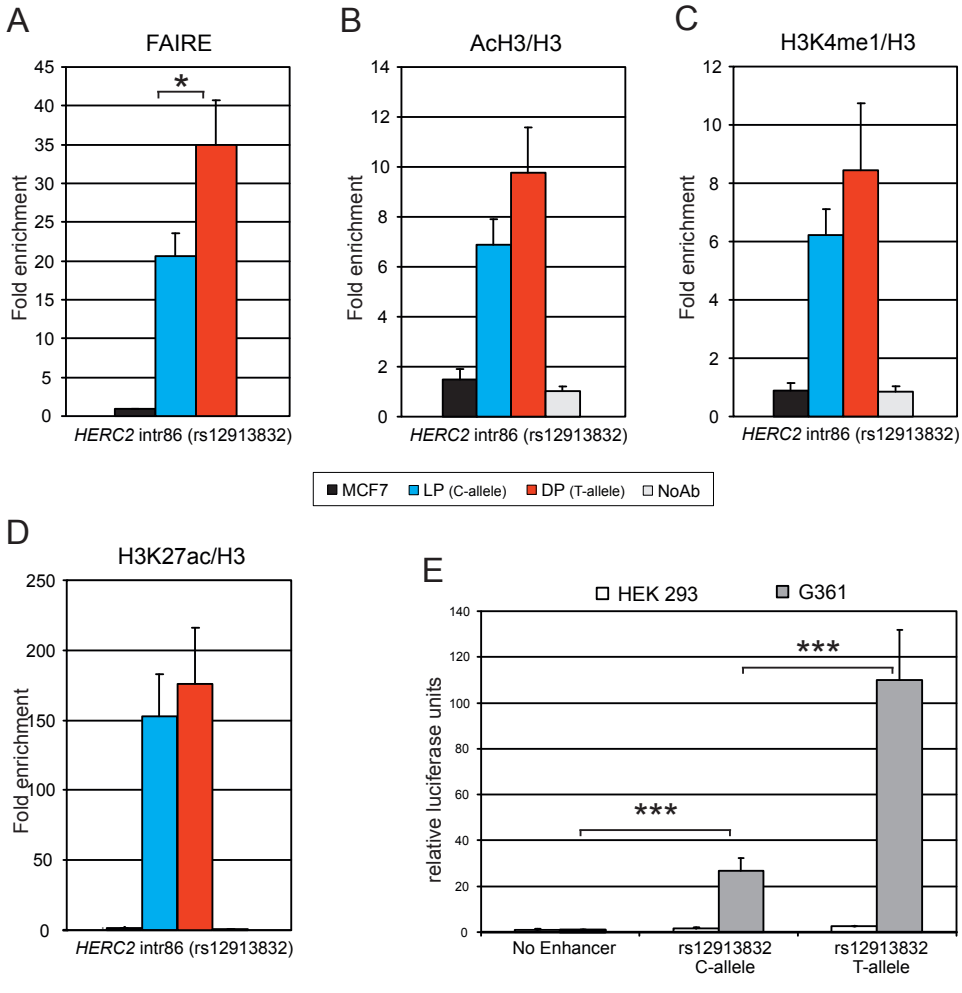


Figure 2. The region directly surrounding *HERC2* rs12913832 acts as a melanocyte-specific enhancer. (A) Formaldehyde assisted identification of regulatory elements (FAIRE) demonstrates low nucleosome occupancy at the rs12913832 region in HEMn-LP and HEMn-DP cells. (B) ChIP-qPCR of Acetylated histone H3 demonstrates active chromatin marks at the rs12913832 region in HEMn-LP and HEMn-DP cells. (C) ChIP-qPCR of histone H3 mono methylated on lysine 4 demonstrates that the rs12913832 region in HEMn-LP and HEMn-DP cells has enhancer potential. (D) ChIP-qPCR of histone H3 acetylated on lysine 27 demonstrates that the enhancer at the rs12913832 region in HEMn-LP and HEMn-DP cells is active. Enrichments displayed are relative to *NDN*. ChIP values for histone H3 marks are normalized to histone H3 occupancy. (E) Luciferase reporter assay demonstrates differential melanocyte enhancer activity for the rs12913832 region. The rs12913832 region from HEMn-LP (C-allele) and HEMn-DP (T-allele) was inserted into a luciferase reporter plasmid and transfected into HEK293 or G361 melanoma cells. Luciferase expression is normalised to *Renilla* luciferase expression. Data are represented as mean ± SEM (*) $p < 0.05$; (***) $p < 0.005$.

reporter vector. Upon transfection of these constructs into Human Embryonic Kidney cells (HEK293), no substantial increase in luciferase expression was observed for the vectors containing the putative enhancer regions, as compared to the empty vector (Fig. 2E). However, robust luciferase expression was observed when these constructs were transfected into G361 melanoma cells, with an approximately 5-fold higher activity for the rs12913832 T-allele than for the rs12913832 C-allele (Fig. 2E; $p < 0.005$). Sequencing of the cloned fragment revealed the presence of another SNP, rs6497271 (Supplemental Fig. S2A), which could theoretically modulate the observed enhancer activity. Analysis of chimeras of our reporter constructs containing different combinations of rs12913832 alleles and rs6497271 alleles, however, demonstrates that the enhancer activity of the *HERC2* rs12913832 region is not modulated by rs6497271 (Supplemental Fig. S2B). These findings also confirmed the previously reported suggestion of Sturm et al. (2008) that this SNP could not explain association to the region due very low minor allele frequency (Sturm et al. 2008). Together, these independent experiments demonstrate that the region surrounding *HERC2* rs12913832 acts as a melanocyte-specific enhancer whose activity depends on the allelic status of rs12913832.

Detection of a differential chromatin loop between the HERC2 rs12913832 enhancer and the OCA2 promoter

Distal enhancers increase the expression of their target genes by directly contacting the gene promoters via chromatin loops, which can be detected using chromosome conformation capture (3C) technology (Palstra 2009). In 3C, formaldehyde is used to trap interactions between chromatin segments which are in close proximity to each other. The cross-linked chromatin is subsequently digested using a restriction enzyme followed by intra-molecular ligation under dilute conditions. The relative abundance of ligation products is determined by qPCR and is proportional to the frequency with which the various restriction fragments interact. We inferred that if the region surrounding *HERC2* rs12913832 acts as a melanocyte-specific enhancer for the *OCA2* gene, as indicated by our data, we should be able to detect a chromatin loop between the *HERC2* rs12913832 region and the *OCA2* gene. Indeed, locus-wide 3C analysis of an EcoRI restriction fragment containing rs12913832 showed, in HEMn-DP cells, an elevated relative cross-linking frequency for an *OCA2* promoter-containing restriction fragment, indicating a chromatin loop between the rs12913832 T-allele and the *OCA2* promoter (Fig. 3A). HEMn-LP cells, however, showed a lower cross-linking frequency of this fragment (Fig. 3A), demonstrating a decreased interaction efficiency between the rs12913832 C-allele and the *OCA2* promoter. In contrast, MCF7 control cells showed a linear chromatin conformation (Fig. 3A).

To further investigate the observed differences in chromatin looping, we aimed for a higher resolution of the 3C experiments. We therefore used a more frequently digesting restriction enzyme (ApoI versus EcoRI) that generates smaller restriction fragments. Analysis of locus-wide cross-linking frequencies of the ~500-bp ApoI fragment containing *HERC2* rs12913832 demonstrates a chromatin loop between rs12913832 and the *OCA2* promoter that is more prominent in HEMn-DP cells (T-allele) than in HEMn-LP cells (C-allele) (Fig. 3B; $p < 0.05$). The difference in chromatin looping between the two *HERC2* rs12913832 alleles and

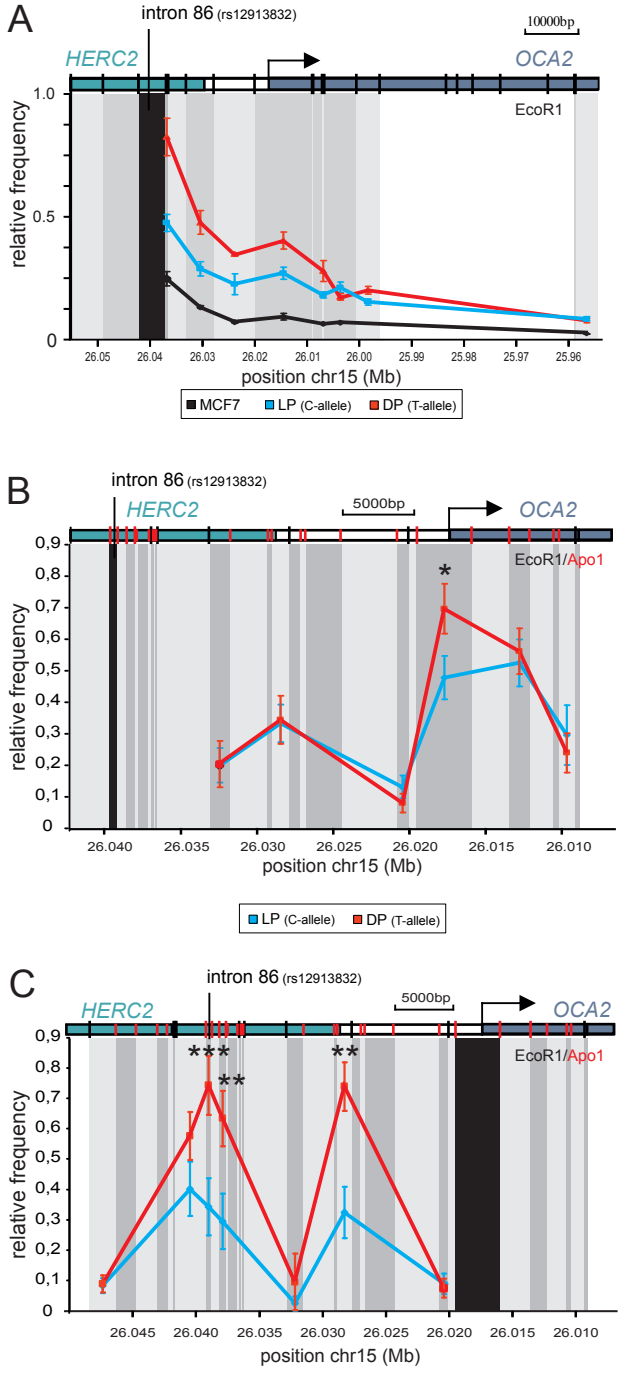


Figure 3. *(legend at bottom of the next page)*

the *OCA2* promoter becomes even clearer in the reciprocal analysis of the ApoI fragment that contains the *OCA2* promoter (Fig. 3C; $p < 0.005$). Locus-wide 3C analysis in MCF7 control cells revealed a more linear chromatin organization (Supplemental Fig. S3A,B), and an interaction between *HERC2* rs12913832 and the *OCA2* promoter is absent (Supplemental Fig. S3C; $p < 0.005$). These data demonstrate that the formation of a chromatin loop between *HERC2* rs12913832 and the *OCA2* promoter is melanocyte-specific and is more efficient when the rs12913832 T-allele is present; this is in line with the higher *OCA2* expression observed in HEMn-DP cells (T-allele) (Fig. 1C).

Interestingly, in both experiments, an additional peak in cross-linking frequencies was detected with a restriction fragment downstream of the *HERC2* gene, suggesting the presence of an additional regulatory element in this region (Fig. 3B,C). This peak is also detected in MCF7 cells although at a lower frequency (Supplemental Fig. S3A,B, Fig. S4A) and therefore probably reflects a constitutive interaction not involved in melanocyte specific *OCA2* regulation. ENCODE data (The ENCODE Project Consortium et al. 2007) obtained from multiple cell lines and our own ChIP-qPCR experiments, indeed, demonstrate the presence of chromatin signatures associated with gene-regulatory elements in the *OCA2-HERC2* intergenic region, thereby confirming this notion (Supplemental Fig. S4B,C).

Detection of an allele-specific chromatin loop between the *HERC2* rs12913832 enhancer and the *OCA2* promoter in combination with the previously observed link between rs12913832 alleles and *OCA2* expression (Cook et al. 2009) demonstrate that the *HERC2* rs12913832 enhancer directly regulates *OCA2* expression.

Figure 3. A chromatin loop is formed between the *HERC2* rs12913832 enhancer region and the *OCA2* promoter. (A-C) Locus-wide cross-linking frequencies observed in MCF7 (black), HEMn-LP (cyan) and HEMn-DP (red) cells. The analyzed region of the human *OCA2-HERC2* locus is depicted on *top* of each graph. The *x*-axis shows the approximate position on chromosome 15 (UCSC browser NCBI36/hg18 assembly; see also Fig. 1A). (Black shading) The position and size of the 'fixed' restriction fragment; (grey shading) position and size of other restriction fragments analyzed. (Black vertical bars in the locus graph) EcoRI sites; (red vertical bars) ApoI sites. The crosslinking frequencies are normalized to the highest interaction within an experiment. (A) Cross-linking frequencies for an EcoRI restriction fragment containing rs12913832 in MCF7 and HEMn cells. In HEMn cells, high cross-linking frequencies are observed for a restriction fragment containing the *OCA2* promoter. (B) Cross-linking frequencies for an ApoI restriction fragment containing rs12913832 in HEMn cells. In HEMn cells, high cross-linking frequencies are observed for a restriction fragment containing the *OCA2* promoter. Cross-linking frequencies between the restriction fragment containing rs12913832 and the restriction fragment containing the *OCA2* promoter are higher for the T-allele (red) than for the C-allele (cyan). (C) Cross-linking frequencies for an ApoI restriction fragment containing the *OCA2* promoter in HEMn cells. High cross-linking frequencies with restriction fragments surrounding the rs12913832 enhancer region are observed in HEMn cells. Cross-linking frequencies between the restriction fragment containing rs12913832 and the restriction fragment containing the *OCA2* promoter are higher for the T-allele (red) than for the C-allele (cyan). Data are represented as mean ± SEM (*) $p < 0.05$; (**) $p < 0.01$; (***) $p < 0.005$.

Analysis of transcription factor binding at the HERC2 rs12913832 enhancer

Sturm et al. (2008) performed an in silico analysis of transcription factor binding sites present in the *HERC2* region surrounding rs12913832. Potential binding sites for HLTF (also known as RUSH or SMARCA3), MITF and LEF1 were identified (Supplemental Fig. S2A), and it was hypothesized that the C-allele of rs12913832 would disrupt the HLTF binding site (Sturm et al. 2008; Sturm and Larsson 2009). Both HLTF and LEF1 have previously been implicated in enhancer function and chromatin looping (Hewetson et al. 2008; Yun et al. 2009), whereas MITF is a key transcription factor in melanocyte development and pigment gene expression (Levy et al. 2006). To experimentally verify whether the *HERC2* rs12913832 alleles are differentially bound by HLTF, we performed ChIP-qPCR experiments. HLTF binding was enriched above background for the rs12913832 T-allele, but not for the C-allele (Fig. 4A; $p < 0.01$), providing experimental evidence that the T-to-C mutation reduces HLTF binding to the rs12913832 region. In order to further investigate the role of HLTF in *OCA2* expression we overexpressed HLTF in HEMn-LP and HEMn-DP cells. In HEMn-LP cells overexpression of HLTF resulted in reduced *OCA2* expression (Fig. 4B). However, *HERC2* expression was also reduced in HEMn-LP and in HEMn-DP cells, but to a similar extend, which therefore suggests that the reduced *OCA2* expression in HEMn-LP cells is a systematic response rather than a functional response to HLTF overexpression (data not shown). In contrast, overexpression of HLTF in HEMn-DP cells resulted in elevated expression levels of *OCA2* (Fig. 4B; $p < 0.05$), demonstrating a positive role of HLTF in the regulation of *OCA2* expression and this regulatory role depends on the T-allele of *HERC2* rs12913832.

Reduced recruitment of HLTF, a chromatin remodeler of the SWI-SNF family (MacKay et al. 2009), to the rs12913832 C-allele may lead to a more closed chromatin conformation that could affect the recruitment of other transcription factors such as MITF and LEF1. Indeed, we found increased binding of MITF and LEF1 to the region encompassing rs12913832 containing the T-allele, in comparison to this region containing the C-allele (Fig. 4C,D; $p < 0.05$). This is in line with a recent study demonstrating that MITF binds near *HERC2* rs12913832 in a melanoma cell line (Strub et al. 2010). Furthermore, in melanoma cells, *OCA2* expression is upregulated upon overexpression of MITF (Hoek et al. 2008). In agreement with this, we found that overexpression of MITF in HEMn-LP cells resulted in increased *OCA2* expression (Fig. 4E; $p < 0.05$) while over-expression of a dominant negative form of MITF (dnMITF) (Levy et al. 2010) resulted in reduced *OCA2* expression (Fig. 4E; $p < 0.05$). Interestingly, overexpression of MITF in HEMn-DP cells failed to increase *OCA2* expression, suggesting that binding of MITF to the rs12913832 enhancer is not limiting in HEMn-DP cells (Fig. 4E). However, overexpression of the dnMITF in HEMn-DP cells resulted in decreased *OCA2* expression (Fig. 4E; $p < 0.005$). Together these data demonstrate that MITF indeed regulates *OCA2* expression. It suggests that MITF preferentially binds to the *HERC2* intron 86 region when it contains the rs12923832 T-allele, which is facilitated by the interaction of HLTF to the rs12913832 region.

Analysis of pigmentation-associated SNPs other than HERC2 rs12913832

It remains possible that linkage with another DNA variant may be responsible for the

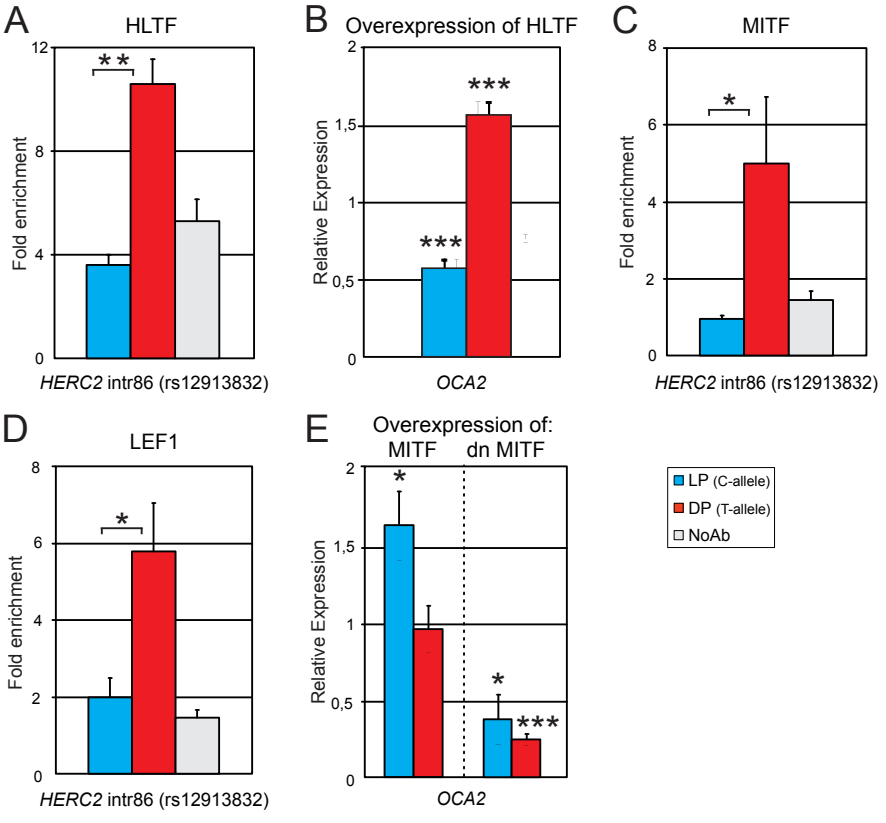


Figure 4. The *HERC2* rs12913832 enhancer is regulated by the transcription factors HLTF, MITF and LEF1. (A) ChIP-qPCR of HLTF at the rs12913832 region in HEMn-LP (C-allele) and HEMn-DP (T-allele) cells. HLTF binding is only observed for the T-allele. (B) Overexpression of HLTF in HEMn-DP cells results in increased *OCA2* expression but not in HEMn-LP cells. (C) ChIP-qPCR of MITF at the *HERC2* rs12913832 region in HEMn-LP (C-allele) and HEMn-DP (T-allele) cells. MITF binding is only observed for the T-allele. (D) ChIP-qPCR of LEF1 at the *HERC2* rs12913832 region in HEMn-LP (C-allele) and HEMn-DP (T-allele) cells. LEF1 binding is only observed for the T-allele. ChIP enrichments displayed are relative to *NDN*. (E) Overexpression of MITF in HEMn-LP cells results in increased *OCA2* expression. This is not observed in HEMn-DP cells. Overexpression of a dominant negative MITF (dnMITF) results in decreased *OCA2* expression in both HEMn-LP and HEMn-DP cells. Expression is relative to *ACTB* expression and nontransfected control cells. Data are represented as mean ± SEM (*) $p < 0.05$; (**) $p < 0.01$; (***) $p < 0.005$).

regulatory effect of a noncoding SNP. To exclude the possibility that pigmentation-linked sequence variants other than *HERC2* rs12913832 might explain our observations, we genotyped all SNPs within the analyzed 3'*HERC2*/5'*OCA2* region that have been linked to pigmentation, including those in linkage disequilibrium (LD; $r^2 > 0.8$) (Supplemental Table S1). This region of ~35kb contains 12 pigmentation-linked SNPs in total. Besides *HERC2* rs12913832, six SNPs have been directly linked to pigmentation traits (Liu et al. 2009), while five additional SNPs are in high LD ($r^2 > 0.8$) with these six SNPs (Supplemental Table S1).

The HEMn-LP and HEMn-DP cells share the same genotype of only one pigmentation-linked SNP, namely, rs7183877, and have different alleles for the other five SNPs, whereas for the linked SNPs, the allelic status for all but one SNP is the same in HEMn-LP and HEMn-DP cells. The LD-SNP rs7497270 is homozygous in the HEMn-LP cells (C) and heterozygous in the HEMn-DP cells (CT). To assess the possible regulatory potential of these additional 11 pigmentation-linked SNPs, we interrogated these sites as well as several control sites using FAIRE analysis in HEMn-LP and HEMn-DP cells, and in the MCF7 control cell line (Fig. 5). As we showed before (Fig. 2A), *HERC2* rs12913832 is highly enriched in HEMn cells as compared to MCF7 cells. The only other site with such a high enrichment is rs7183877. This site shares the C-allele in HEMn-LP and HEMn-DP cells and is located only 115 bp away from *HERC2* rs12913832. This close physical location explains its high enrichment in the FAIRE assay, because FAIRE analysis has a resolution in the range of 0.5-1.5 kb. From the remaining investigated sites, only those in the promoter region and in the first intron of *OCA2* have an appreciable although small enrichment in the FAIRE assay (Fig. 5). Irrespective of their genotypes this enrichment appears to be higher in the HEMn-DP cells than in the HEMN-LP cells probably reflecting the higher transcriptional status of *OCA2* in the HEMN-DP cells (Fig. 1C,D). We therefore conclude that none of the SNPs that can be linked to pigmentation within the ~35kb *HERC2*/*OCA2* region analyzed other than rs12913832 has regulatory potential.

Furthermore, of the 12 pigmentation-linked SNPs only three – rs12913832, rs7164220 and rs4778241 – are significantly conserved among vertebrate species, which might suggest functionality (Supplemental Table S1). Both HEMn-LP and HEMn-DP cells share the C-allele for rs7164220. However, HEMn-LP and HEMn-DP cells differ in alleles for rs4778241 (C and A), and therefore it remains possible that the observed allele-specific chromatin loop between *HERC2* rs12913832 and the *OCA2* promoter is caused by this SNP. To investigate this, we made use of a commercially available cell line derived from an intermediate pigmented donor (HEMn-MP) (Cascade Biologics, Invitrogen). This cell line is homozygous for the C-allele of *HERC2* rs12913832 but heterozygous for rs4778241 (C and A) (Supplemental Table S1). The presence of the rs4778241 C-allele generates an MfeI restriction site, which allowed us to design an allele-specific 3C assay (Fig. 6A). We used this assay to test the hypothesis that rs4778241, instead of *HERC2* rs12913832, is responsible for the observed allele-specific chromatin loop between *HERC2* rs12913832 and the *OCA2* promoter. If this hypothesis holds true, we would expect to find an increased interaction of the ApoI restriction fragment containing *HERC2* rs12913832 with the ApoI restriction fragment containing the A-allele of rs4778241 (also present in HEMn-DP cells). Instead, however, we detected an interaction between the rs12913832 ApoI fragment and both the ApoI fragments containing either the A-allele or C-allele of rs4778241 (Fig. 6B,C), and there seemed to be a preference (although not statistically significant) for the rs4778241 C-allele, which is also present in HEMn-LP cells. The ApoI restriction fragment containing *HERC2* rs12913832 does not contain any sequence variation except for *HERC2* rs12913832 (supplemental figure 2A), enforcing the notion that the observed differential chromatin looping is solely dependent on the *HERC2* rs12913832 allele.

Together these observations make it highly improbable that the observed allele-specific chromatin loop between *HERC2* rs12913832 and the *OCA2* promoter is caused by a

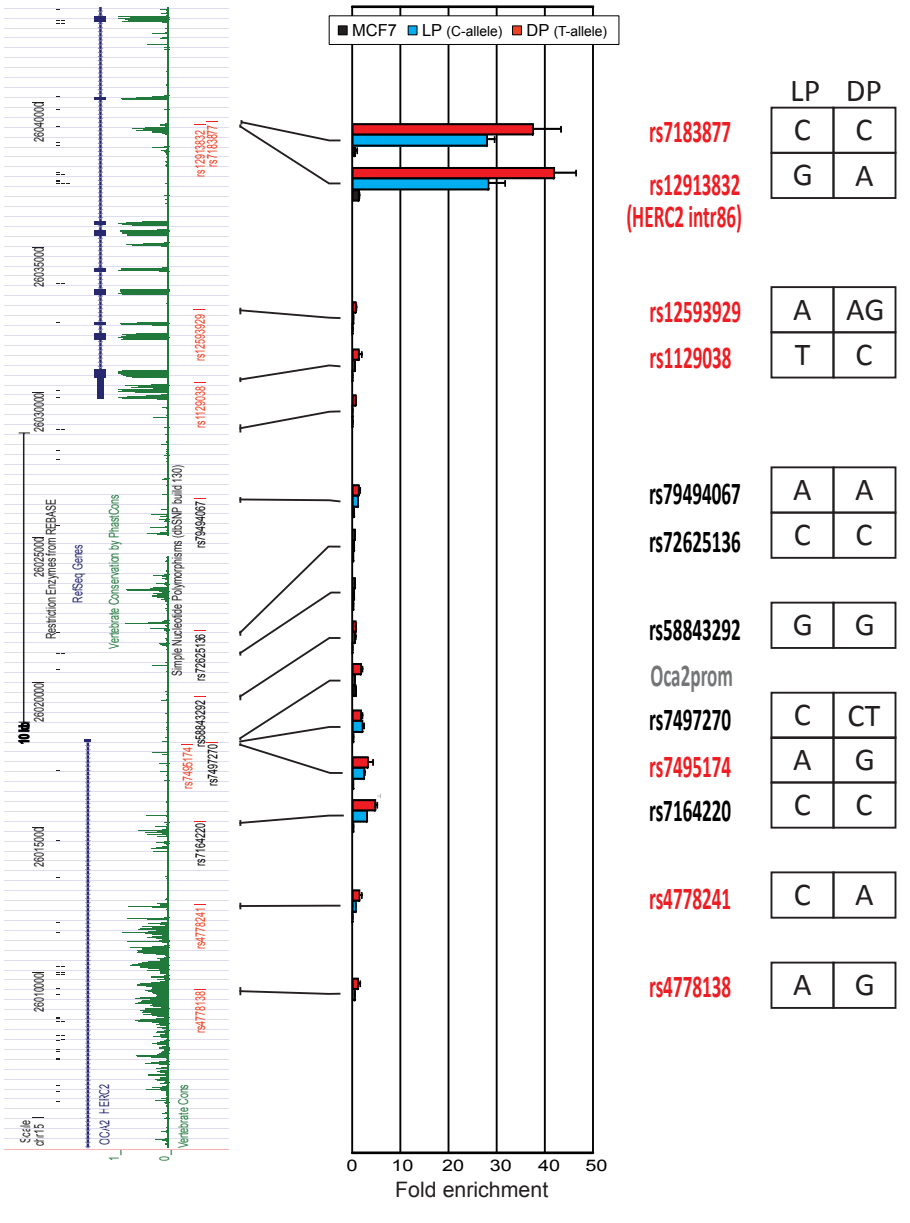


Figure 5. FAIRE analysis of pigmentation-associated SNPs other than *HERC2* rs12913832 present within the 3'*HERC2*/5'*OCA2* region does not reveal additional regulatory elements. (*Left* side) Tracks from the UCSC browser (NCBI36/hg18 assembly; http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg18) of the investigated 3'*HERC2*/5'*OCA2* region. Pigmentation-associated SNPs are indicated (red); linked SNPs ($r^2 > 0.8$) (black). The approximate location of the analyzed PCR amplicons is indicated. (*Right* side) The genotype of each SNP for HEMn-LP and HEMn-DP. Enrichments displayed are relative to *NDN*. Data are represented as mean ± SEM.

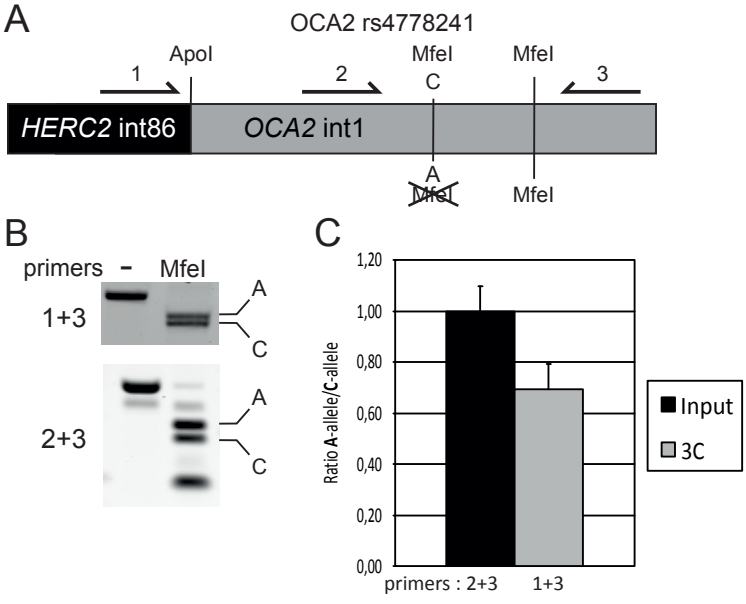


Figure 6. The chromatin loop between the *HERC2* rs12913832 enhancer region and the *OCA2* promoter is not caused by allelic differences in the 5' region of *OCA2*. (A) Schematic overview of the allele-specific 3C assay. The allele-specific interaction between an ApoI fragment containing *HERC2* rs12913832 and an ApoI fragment containing *OCA2* rs4778241 is investigated. The presence of the C-allele of rs4778241 generates an additional MfeI restriction site. Primers 1 and 3 are used to detect the 3C product while primers 2 and 3 are used to normalize the ratio of the A-allele over the C-allele. Ratios are determined by MfeI digestion of PCR products. The additional MfeI site is used to monitor the completeness of digestion. (B) Example of gel images of a representative MfeI digestion of PCR products. In the digested lane, the *top* band represents the A-allele and the *bottom* band the C-allele of rs4778241. (C) Quantification of multiple gel images as shown in *B*. The ratio of the bands generated by primer pair 2 + 3 after MfeI digestion is set to 1. Data are represented as mean ± SEM.

SNP other than *HERC2* rs12913832 located in the 3'*HERC2*/5'*OCA2* region.

Discussion

Overall, our study provides extensive experimental evidence that the genomic region directly surrounding *HERC2* rs12913832 functions as a human melanocyte-specific enhancer element that influences *OCA2* gene expression. In particular, we demonstrate that the *HERC2* rs12913832 enhancer communicates with the *OCA2* promoter via a long-range chromatin loop, and enhancer activity is mediated by the transcription factors HLTF, LEF1 and MITF. The *HERC2* rs12913832 T-allele robustly recruits these transcription factors, which, in combination with increased looping to the *OCA2* promoter, facilitates *OCA2* expression. We propose that this leads to enhanced melanin production and the dark pigmentation phenotype. In contrast, the *HERC2* rs12913832 C-allele mutates an HLTF binding site and diminishes HLTF recruitment, which correlates with decreased MITF and LEF1 recruitment

and reduced chromatin-loop formation. This results in reduced *OCA2* expression and the light pigmentation phenotype.

Linkage with an unknown causal DNA variant may be indicated as a plausible explanation for the regulatory effect of a noncoding SNP. We consider this hypothesis however unlikely for the following reasons: Our data clearly demonstrate that the region around *HERC2* rs12923832 displays chromatin features associated with enhancers; it acts as a melanocyte-specific enhancer in luciferase reporter assays; it communicates with the *OCA2* promoter region via a chromatin loop; and the transcription factors HLTF, MITF and LEF interact with the *HERC2* rs12913832 region in an allele-specific manner. Together, this demonstrates that the *HERC2* rs12913832 region constitutes a genuine regulatory element for the *OCA2* gene. Moreover, other pigmentation SNPs within the investigated region do not display chromatin features consistent with enhancer function and do not influence chromatin looping between *HERC2* rs12923832 and the *OCA2* promoter region. Considering our data in combination with previously published indirect evidence such as (1) the gene function of *OCA2* being one of the major pigmentation genes that affects the amount and quality of melanin in melanocytes (Brilliant 2001); (2) several *OCA2* mutations results in albinism (Rinchik et al. 1993; Brilliant 2001); and (3) a strong correlation between *HERC2* rs12913832, *OCA2* expression levels, and melanocytic melanin content (Cook et al. 2009), it is highly likely that *HERC2* rs12913832 is the major causal variant for the regulation of *OCA2* transcription and hence pigmentation. Our data, however, do not exclude the possibility that additional regulatory regions for the *OCA2* gene are present within or even outside the *HERC2*/*OCA2* region, but if existing at all, their effect is expected to be rather minor.

Previously, two different models proposed that the region around *HERC2* rs12913832 acts as a regulatory element that influences *OCA2* expression. Eiberg et al. (2008) suggested that the *HERC2* rs12913832 region functions as a silencer for *OCA2* expression, while Sturm and colleagues (Sturm et al. 2008; Sturm and Larsson 2009) suggested that the *HERC2* rs12913832 region functions as an enhancer. Clearly, our study provides solid experimental evidence to support the model as proposed by Sturm and colleagues (Sturm et al. 2008; Sturm and Larsson 2009) and further extends it, since we demonstrate that the region around *HERC2* rs12913832 displays multiple features of an enhancer element and forms an allele-specific chromatin loop with the *OCA2* promoter. Since both HLTF and LEF1 have previously been shown to be involved in chromatin looping, we find it plausible that the increased binding of these factors to the *HERC2* rs12913832 T-allele is responsible for the observed increased looping of the *HERC2* rs12913832 region to the *OCA2* promoter (Hewetson and Chilton 2008; Yun et al. 2009). The model proposed by Sturm and colleagues suggests that binding of MITF and LEF1 to the *HERC2* intron 86 region fully depends on HLTF interaction with *HERC2* rs12913832. However, based on our data, this part of the model requires some refinement. Using ChIP assays we only detected MITF and LEF1 binding when the T-allele of *HERC2* rs12913832 is present, but due to the inefficiency of ChIP assays, it is plausible that we missed low-level binding of these factors to the rs12913832 C-allele. Our overexpression studies demonstrate that the locus containing the C-allele of rs12913832 is responsive to MITF levels but irresponsive to HLTF levels. In contrast, the locus containing the rs12913832 T-allele is responsive to HLTF overexpression and to expression of the dnMITF but not to increased MITF levels. This suggests that MITF binding to the *HERC2* rs12913832 region in

HEMn cells is in an equilibrium which can be modulated by varying MITF levels and HLTF binding. Alternatively, other MITF responsive *OCA2* regulatory elements might be present.

For some disease-associated SNPs such as the cancer risk variants rs6983267 (Pomerantz et al. 2009; Ahmadiyeh et al. 2010; Wright et al. 2010) and rs11986220 (Jia et al. 2009), or DNA variants associated with coronary artery disease (Harismendy et al. 2010), it has been reported that they constitute transcription regulatory elements for distal genes, and the presence of a chromatin loop between the putative regulatory element and the promoter of the regulated gene has been demonstrated. However, none of these studies report allelic differences in chromatin-loop formation. One study (Wright et al. 2010) that specifically investigated potential allelic differences in chromatin-loop formation between rs6983267 and the MYC gene in colon cancer cells failed to demonstrate such an allelic difference. In contrast, our study provides the key mechanistic insight that allele-dependent differences in chromatin-loop formation (i.e., structural differences in the folding of gene loci) functionally contribute to differences in allelic gene expression. Moreover, the predicted risk associated with above-mentioned rare disease phenotypes is modest, whereas *HERC2* rs12913832 has strong predictive power for human pigmentation phenotypes (Liu et al. 2009, 2010; Valenzuela et al. 2010; Branicki et al. 2011; Spichenok et al. 2010;). Therefore, our study demonstrates that genetic variation in gene regulatory elements can have a strong influence on common human phenotypic traits, which extends previous knowledge from rare phenotypes and puts it into a new and more general perspective.

We anticipate that many phenotypic traits, pathological and non-pathological, are modulated by DNA variants in distally-located regulatory regions. In this study we demonstrate that it is feasible to annotate regulatory function to a noncoding SNP and identify its target gene, using an approach that combines FAIRE, ChIP and 3C. All these techniques can be used at a genome-wide scale to study the epigenome landscape. Such a genome-wide approach combined with GWAS data allows to assign function to many more noncoding SNPs in future studies that aim to unveil the functional basis of genetically determined phenotypic variation.

Methods

Cell culture

HEMn-LP (C-0025C; Lot #200708522; Cascade Biologics, Invitrogen) HEMn-MP (C-1025C ; Lot# 070417901; Cascade Biologics, Invitrogen), and HEMn-DP (C-2025C, Lot# 6C0474; Cascade Biologics, Invitrogen) were grown in Medium 254 supplemented with HMGS according to the manufacturer's instruction (Cascade Biologics, Invitrogen). G361, HEK293 and MCF7 cells were cultured in DMEM/10% FCS at 37°C/5% CO_2. HLTF was overexpressed using a pCDNA5 FRT/TO/FLAG SMARCA3 (MacKay et al. 2009). Transfection was done using Lipofectamine LTX transfection reagent (Invitrogen) according to the manufacturer's instructions. MITF and dnMITF were overexpressed using adeno viral constructs (a kind gift from Dr. D. Fisher).

Genotyping

Pigmentation SNPs including *HERC2* rs12913832 were genotyped with the iPLEX (Sequenom) multiplex reactions as described by Liu et al (2009) using genomic DNA samples derived from the different cell lines. Genotypes for rs12913832 were confirmed by direct sequencing of PCR fragments containing rs12913832. Genotypes of rs7164220, rs7497270, rs58843292, rs72625136 and rs79494067 were determined by restriction digestion of PCR fragments. To infer the geographic origin/genetic ancestry of the cell-line donors, 24 autosomal SNPs sensitive to detect continental ancestry were genotyped via two SNaPshot (Applied Biosystems) multiplex reactions as described elsewhere (Lao et al. 2010). The continental ancestry of the samples was recovered by performing a STRUCTURE analysis (Pritchard et al. 2000) in which data from the HGDP-CEPH panel served as reference as described elsewhere (Lao et al. 2010).

Chromatin Immuno-Precipitation (ChIP)

ChIP was performed as described in the Millipore protocol (http://www.millipore.com/userguides/tech1/mcproto407), except that samples were cross-linked with 2% formaldehyde for 10 minutes at room temperature. To remove melanin, DNA was column purified before PCR (OneStep™ PCR Inhibitor Removal Kit, Zymo Research). Quantitative real-time PCR (CFX96TM Real Time System, Bio-Rad) was performed using SYBR Green (Sigma) and Platinum *Taq* DNA Polymerase (Invitrogen) under the following cycling conditions: 10 min at 50°C, 5 min at 95°C, 45 cycles of 10 sec at 95°C, 30 sec at 60°C, and followed by a melting curve analysis. Enrichment was calculated relative to necdin (*NDN*), and values were normalized to input measurements. Antibodies used: RNA Polymerase II (POLR2A, N-20; sc-899), LEF1 (H-70; sc-28687X), MITF (C-17; sc-11002X) from Santa Cruz Biotechnology, Ac-H3 (#06-599) from Millipore, PanH3 (#ab1791), acetylated Histone H3 K27 (#4729ab) and mono-methyl Histone H3 K4 (#ab8895-50) from Abcam, HLTF (PAB12415) from Abnova. Primer sequences are listed in Supplemental Table S2.

Formaldehyde-Assisted Identification of Regulatory Elements (FAIRE)

FAIRE was performed as described before (Giresi et al. 2007), except that selected genomic sites were analyzed by quantitative real-time PCR (see ChIP). To remove melanin, DNA was column purified before PCR (OneStep™ PCR Inhibitor Removal Kit, Zymo research). Primer sequences are listed in Supplemental Table S2.

Transcription analysis

Total cellular RNA was isolated from the different cell lines with TriPure Isolation Reagent according to the manufacturer's instructions (Roche Diagnostics) or with the RNeasy Plus Mini kit (Qiagen) following the manufacturer's instructions. To remove PCR-inhibiting substances such as melanin, RNA samples were column purified (OneStep™ PCR Inhibitor Removal Kit, Zymo Research Corporation) following the manufacturer's instructions. Subsequent DNase I digestion was performed with Ambion's Turbo DNA-free kit (Applied Biosystems) according to the manufacturer's protocol. The reverse-transcriptase (RT) reaction was performed using RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas GmbH) according to the manufacturer's instructions. Quantitative real-time PCR reactions for gene expression analysis were performed on a LightCycler® 480 Realtime PCR System (Roche Diagnostics Nederland B.V.) with the Lightcycler 480 SYBR Green I master mix using the following parameters: initial denaturation for 10 min at 95°C, followed by 45 cycles for 10 sec at 95°C, for 30 sec at 60°C, and for 1 sec at 72°C, followed by a melting curve analysis and a final cooling step to 40°C. LightCycler 480 System Software v1.5 (Roche) was used to analyze the qPCR data. The reference gene *ACTB* was used to normalize the amplification signal between the samples of different cell lines, differences in treatment, and amount of input cDNA. Primer sequences are listed in Supplemental Table S2.

Luciferase assays

A 1450-bp fragment surrounding rs12913832 was PCR-amplified from genomic DNA obtained from HEMn-LP (C-allele) or HEMn-DP (T-allele) using the Expand Long Template PCR kit (Roche). The PCR fragment was digested with HindIII and PstI, generating a 750-bp fragment which was subcloned in pBluescript. Sequencing verified the presence of the correct rs12913832 allele for each construct. Correct clones were digested with BamHI and Asp718I and cloned into the BglII/Asp718I sites of a modified pGL3-promoter vector (in which the SV40 promoter and the SV40 3'UTR were replaced by an HSP promoter and an HSP 3'UTR; Promega). Constructs were transfected into HEK293 or G361 melanoma cells using Lipofectamine 2000, and luciferase expression was normalized to *Renilla* luciferase expression. Chimerical luciferase reporter constructs containing different combinations rs12913832 and rs6497271 alleles were generated by digesting HEMn-LP (C-allele) and HEMn-DP (T-allele) containing luciferase reporter constructs with XmnI and religating after swapping the fragments. Constructs were sequence-verified.

Chromosome Conformation Capture (3C) analysis

Chromosome Conformation Capture (3C) analysis was performed essentially as described (Palstra et al. 2003; Hagege et al. 2007) using EcoRI or ApoI as the restriction enzyme. To remove melanin, DNA was column purified before PCR (OneStep™ PCR Inhibitor Removal Kit, Zymo research). Quantitative real-time PCR (CFX96TM Real Time System, Bio-Rad) was performed using iTaq SYBR Green Supermix with ROX (Bio-Rad), under the following cycling conditions: 2 min at 50 °C, 10 min at 95 °C, 45 cycles of 15 sec at 95 °C, 1 min at 60 °C, followed by a melting curve analysis. Cross-linking frequencies between samples were normalized using primers in the nonexpressed human beta hemoglobin LCR and *HBE1* (Palstra et al. 2003). A random template was generated as described previously (Palstra et al. 2003) using BAC RP11-1365A12 for the *OCA2*/*HERC2* locus. For the human beta hemoglobin locus, we used a 185-kb PAC (Palstra et al. 2003). Primer sequences are listed in Supplemental Table S2.

Statistical analysis

Statistical calculations were performed in Excel, or using R software (for the data obtained in the ChIP assays AcH3/H3; H3K4Me1/H3 and H3K27Ac/H3). All *p*-values were calculated with *t*-tests.

Acknowledgments

We thank R.A. Poot for support in initiating this project and for useful comments on the manuscript. We also like to thank F. Grosveld for providing infrastructural support and for useful comments on the manuscript. R. Stadhouders and E. Soler are acknowledged for discussions, J. van Haren for providing the G361 melanoma cell line, D. Fisher (Dana-Farber Cancer Institute) for generously providing MITF constructs, K. van Duijn, M. Vermeulen and Y. Choi for help with genotyping, K. van Duijn and O. Lao for help in genetic ancestry inferences, O. Lao for help with statistical analyses, and S. Walsh for valuable comments on the manuscript. R.J.P. was supported by the Netherlands Organisation for Scientific Research (NWO) grant number 700.57.408, and the Erasmus MC. M.K. was supported in part by the Netherlands Forensic Institute (NFI), the Erasmus MC, and a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

References

Ahmadiyeh, N., M.M. Pomerantz, C. Grisanzio, P. Herman, L. Jia, V. Almendro, H.H. He, M. Brown, X.S. Liu, M. Davis et al. 2010. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci* **107:** 9742-9746.

Branicki, W., U. Brudnik, and A. Wojas-Pelc. 2009. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. *Ann Hum Genet* **73:** 160-170.

Branicki, W., F. Liu, K. van Duijn, J. Draus-Barini, E. Pospiech, S. Walsh, T. Kupiec, A. Wojas-Pelc, and M. Kayser. 2011. Model-based prediction of human hair color using DNA variants. *Hum Genet* **129:** 443-454.

Brilliant, M.H. 2001. The mouse p (pink-eyed dilution) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH. *Pigment Cell Res* **14:** 86-93.

Coetzee, G.A., L. Jia, B. Frenkel, B.E. Henderson, A. Tanay, C.A. Haiman, and M.L. Freedman. 2010. A systematic approach to understand the functional consequences of non-protein coding risk regions. *Cell Cycle* **9:** 256-259.

Cook, A.L., W. Chen, A.E. Thurber, D.J. Smit, A.G. Smith, T.G. Bladen, D.L. Brown, D.L. Duffy, L. Pastorino, G. Bianchi-Scarra et al. 2009. Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. *J Invest Dermatol* **129:** 392-405.

Creyghton, M.P., A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107:** 21931-21936.

Eiberg, H., J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K.W. Kjaer, and L. Hansen. 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 Genet* **123:** 177-187.

The ENCODE Project Consortium E. Birney J.A. Stamatoyannopoulos A. Dutta R. Guigo T.R. Gingeras E.H. Margulies Z. Weng M. Snyder E.T. Dermitzakis et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799-816.

Freedman, M.L., A.N. Monteiro, S.A. Gayther, G.A. Coetzee, A. Risch, C. Plass, G. Casey, M. De Biasi, C. Carlson, D. Duggan et al. 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* **43:** 513-518.

Giresi, P.G., J. Kim, R.M. McDaniell, V.R. Iyer, and J.D. Lieb. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17:** 877-885.

Hagege, H., P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat, and T. Forne. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *NatProtoc* **2:** 1722-1733.

Han, J., P. Kraft, H. Nan, Q. Guo, C. Chen, A. Qureshi, S.E. Hankinson, F.B. Hu, D.L. Duffy, Z.Z. Zhao et al. 2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet* **4:** e1000074. doi:10.1371/journal.pgen.1000074

Harismendy, O., D. Notani, X. Song, N.G. Rahim, B. Tanasa, N. Heintzman, B. Ren, X.D. Fu, E.J. Topol, M.G. Rosenfeld et al. 2010. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* **470:** 264-268.

Heintzman, N.D., R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39:** 311-318.

Hewetson, A. and B.S. Chilton. 2008. Progesterone-dependent deoxyribonucleic acid looping between RUSH/SMARCA3 and Egr-1 mediates repression by c-Rel. *Mol Endocrinol* **22:** 813-822.

Hoek, K.S., N.C. Schlegel, O.M. Eichhoff, D.S. Widmer, C. Praetorius, S.O. Einarsson, S. Valgeirsdottir, K. Bergsteinsdottir, A. Schepsky, R. Dummer et al. 2008. Novel MITF targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res* **21:** 665-676.

Jia, L., G. Landan, M. Pomerantz, R. Jaschek, P. Herman, D. Reich, C. Yan, O. Khalid, P. Kantoff, W. Oh et al. 2009. Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet* **5:** e1000597. doi:10.1371/journal.pgen.1000597

Lao, O., P.M. Vallone, M.D. Coble, T.M. Diegoli, M. van Oven, K.J. van der Gaag, J. Pijpe, P. de Knijff, and M. Kayser. 2010. Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and mitochondrial DNA. *Hum Mutat* **31:** E1875-1893.

Levy, C., M. Khaled, and D.E. Fisher. 2006. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med* **12:** 406-414.

Levy, C., M. Khaled, K.C. Robinson, R.A. Veguilla, P.H. Chen, S. Yokoyama, E. Makino, J. Lu, L. Larue, F. Beermann et al. 2010. Lineage-specific transcriptional regulation of DICER by MITF in melanocytes. *Cell* **141:** 994-1005.

Liu, F., K. van Duijn, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, A.C. Janssens, and M. Kayser. 2009. Eye color and the prediction of complex phenotypes from genotypes. *Curr Biol* **19:** R192-193.

Liu, F., A. Wollstein, P.G. Hysi, G.A. Ankra-Badu, T.D. Spector, D. Park, G. Zhu, M. Larsson, D.L. Duffy, G.W. Montgomery et al. 2010. Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet* **6:** e1000934. doi:10.1371/journal.pgen.1000934

MacKay, C., R. Toth, and J. Rouse. 2009. Biochemical characterisation of the SWI/SNF family member HLTF. *Biochem Biophys Res Commun* **390:** 187-191.

Manolio, T.A. 2010. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* **363:** 166-176.

Palstra, R. 2009. Close encounters of the 3C kind: long-range chromatin interactions and transcriptional regulation. *Briefings in Functional Genomics and Proteomics* **8:** 297-309

Palstra, R.J., B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. de Laat. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35:** 190-194.

Pomerantz, M.M., N. Ahmadiyeh, L. Jia, P. Herman, M.P. Verzi, H. Doddapaneni, C.A. Beckwith, J.A. Chan, A. Hills, M. Davis et al. 2009. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* **41:** 882-884.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure usingmultilocus genotype data. *Genetics* **155:** 945-959.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S.A. Brugmann, R.A. Flynn, and J. Wysocka. 2010. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470:** 279-283.

Rinchik, E.M., S.J. Bultman, B. Horsthemke, S.T. Lee, K.M. Strunk, R.A. Spritz, K.M. Avidano, M.T. Jong, and R.D. Nicholls. 1993. A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. *Nature* **361:** 72-76.

Spichenok, O., Z.M. Budimlija, A.A. Mitchell, A. Jenny, L. Kovacevic, D. Marjanovic, T. Caragine, M. Prinz, and E. Wurmbach. 2010. Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci Int Genet* **5:** 472-478.

Strub, T., S. Giuliano, T. Ye, C. Bonet, C. Keime, D. Kobi, S. Le Gras, M. Cormont, R. Ballotti, C. Bertolotto et al. 2010. Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30:** 2319-2332.

Sturm, R.A. 2009. Molecular genetics of human pigmentation diversity. *Hum Mol Genet* **18:** R9-17.

Sturm, R.A. and M. Larsson. 2009. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res* **22:** 544-562.

Sturm, R.A., D.L. Duffy, Z.Z. Zhao, F.P. Leite, M.S. Stark, N.K. Hayward, N.G. Martin, and G.W. Montgomery. 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 Genet* **82:** 424-431.

Valenzuela, R.K., M.S. Henderson, M.H. Walsh, N.A. Garrison, J.T. Kelch, O. Cohen-Barak, D.T. Erickson, F. John Meaney, J. Bruce Walsh, K.C. Cheng et al. 2010. Predicting phenotype from genotype: normal pigmentation. *J Forensic Sci* **55:** 315-322.

Wright, J.B., S.J. Brown, and M.D. Cole. 2010. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol* **30:** 1411-1420.

Yun, K., J.S. So, A. Jash, and S.H. Im. 2009. Lymphoid enhancer binding factor 1 regulates transcription through gene looping. *J Immunol* **183:** 5129-5137.

Supplemental figures and tables

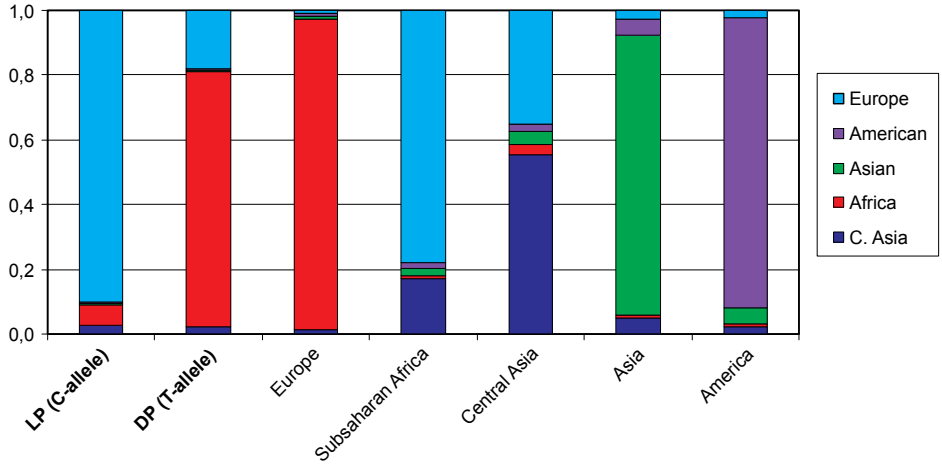


Figure S1 Characterization of HEMn cells. Genetic ancestry of HEMn-LP and HEMn-DP estimated by STRUCTURE using 24 autosomal ancestry-sensitive markers (ASMs) shows a European origin for HEMn-LP and an African origin for HEMn-DP, confirming a light skin colour and dark skin colour of the donors.

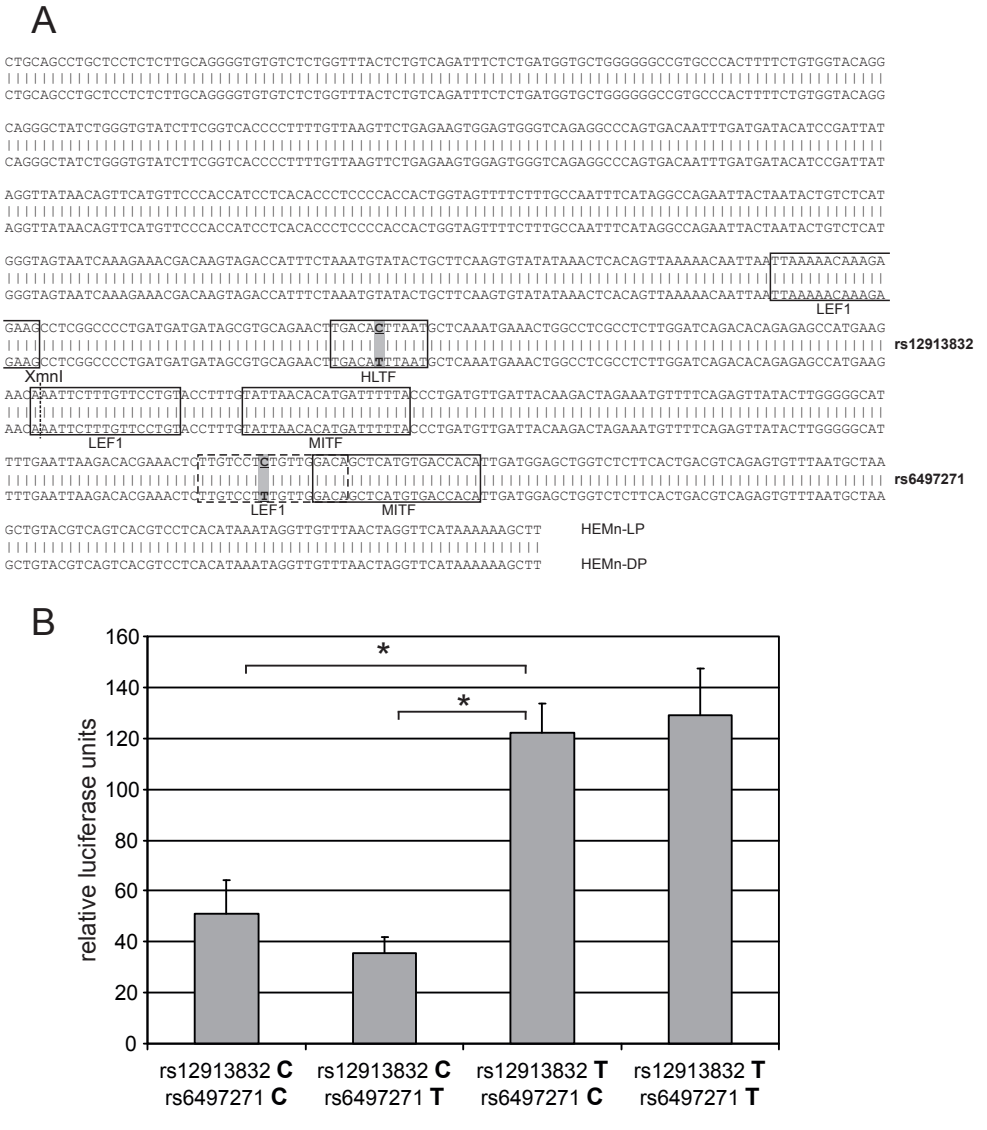


Figure S2 rs6497271 does not influence the activity of the rs12913832 enhancer region. (A) Sequence alignment of the cloned rs12913832 enhancer region from HEMn-LP (top) and HEMn-DP cells (bottom). Positions of rs12913832 and rs6497271 are indicated in bold type and grey shading. The positions of putative binding sites for HLTF, MITF and LEF1 are indicated by boxes (Sturm et al., 2008). The position of the XmnI restriction site used to produce the chimerical luciferase constructs is indicated by a dotted line. (B) Luciferase reporter assay demonstrates that rs6497271 doesn't influence the enhancer activity of the rs12913832 region. Chimerical luciferase reporter plasmids containing the C or T allele of rs12913832 region in combination with the C or T allele of rs6497271 were transfected into G361 melanoma cells and luciferase activity was determined. Luciferase expression is normalised to renilla luciferase expression. Data are represented as mean ± SEM (* = p<0.05).

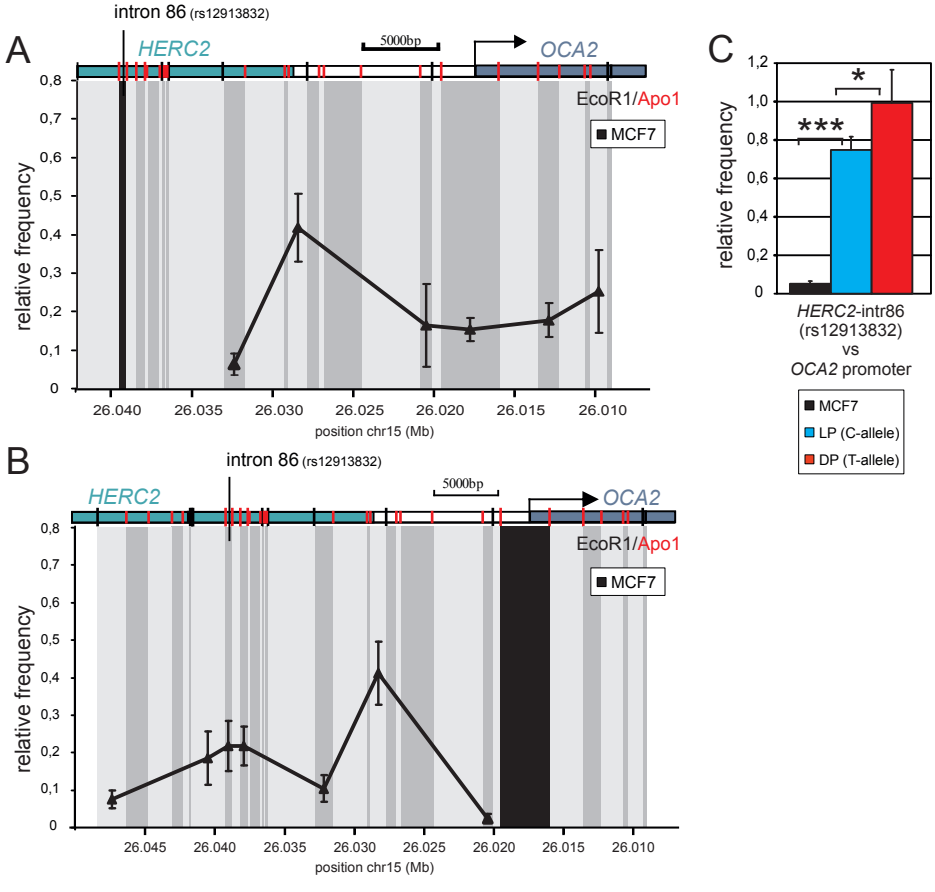


Figure S3 No chromatin loop is present between the *OCA2* promoter and the *HERC2* rs12913832 enhancer region in MCF7 control cells. (A-B) Locus wide cross-linking frequencies observed in MCF7 control cells. The analyzed region of the human *OCA2-HERC2* locus is depicted on top of each graph. X-axis shows the approximate position on chromosome 15 (UCSC browser NCBI36/hg18 assembly; NCBI36/hg18 assembly; http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg18). Black shading shows the position and size of the 'fixed' restriction fragment. Grey shading indicates position and size of other restriction fragments analyzed. Black vertical bars in the locus graph indicate EcoRI sites, red vertical bars ApoI sites. Cross-linking frequencies are normalized to the highest interaction within an experiment. (A) Cross-linking frequencies for an ApoI restriction fragment containing *HERC2* rs12913832 in MCF7 cells. Relatively low cross-linking frequencies with restriction fragments containing the *OCA2* promoter are observed in MCF7 control cells. (B) Cross-linking frequencies for an ApoI restriction fragment containing the *OCA2* promoter in MCF7 cells. Relatively low cross-linking frequencies with restriction fragments containing rs12913832 are observed in MCF7 control cells. (C) Comparison of cross-linking frequencies between an ApoI restriction fragment containing rs12913832 and an ApoI restriction fragment containing the *OCA2* promoter in MCF7 and HEMn cells. Highest interaction is set to 1. Data are represented as mean ± SEM (* = $p<0.05$; *** = $p < 0.005$).

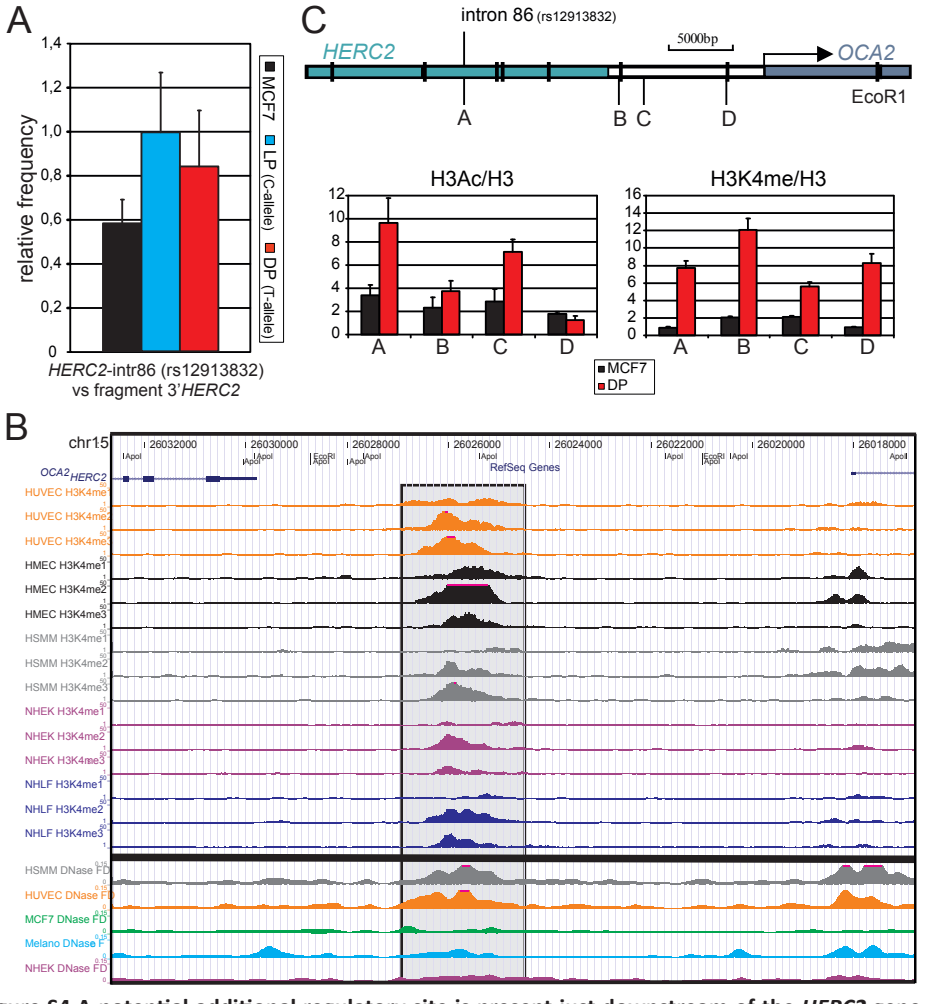


Figure S4 A potential additional regulatory site is present just downstream of the *HERC2* gene. (A) Comparison of crosslinking frequencies between an ApoI restriction fragment containing rs12913832 and an ApoI restriction fragment containing the region just downstream of *HERC2* in MCF7 and HEMn cells indicates a comparable interaction frequency in all cell lines. Highest interaction is set to 1. (B) Tracks from the UCSC browser (NCBI36/hg18 assembly; http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg18) displaying several chromatin marks associated with regulatory elements (Top panel) and DNase sensitivity (bottom panel) in the human *OCA2-HERC2* intergenic region as determined by the ENCODE consortium are shown. A region of "enhancer" chromatin marks (boxed) is detected in the region between *HERC2* and *OCA2*. (C) ChIP-qPCR of Acetylated histone H3 and histone H3 mono methylated on lysine 4 demonstrates active chromatin marks in the *OCA2-HERC2* intergenic region in MCF7 and HEMn-DP cells. A schematic overview of the locus is depicted on top of the figure. Approximate position on chromosome 15 of the PCR amplicons is indicated with letters A = 26.0392Mb; B = 26.0287Mb; C = 26.0267Mb; D = 26.0210Mb. ChIP enrichment displayed are relative to NDN. Data are represented as ± SEM.

Supplemental Table S1. Pigmentation SNPs in the 3' *HERC2* / 5' *OCA2* region.

	LP	MP	DP		phastCons44way
HERC2_rs1129038CT	T	T	C		-1.02779
HERC2_rs12593929AG	A	A	AG		-0.247969
HERC2_rs12913832AG	**G**	**G**	**A**		**1.20791**
HERC2_rs7183877AC	C	C	C		-0.728937
OCA2_rs4778138AG	A	GA	G		-0.0823897
OCA2_rs4778241AC	C	CA	A		0.64885
OCA2_rs7495174AG	A	AG	G		-0.897596
OCA2_rs7164220CA	C	nd	C		0.369879
OCA2_rs7497270CT	C	nd	CT		-0.0969921
Int_rs58843292GA	G	nd	G		-0.441677
Int_rs72625136CT	C	nd	C		-1.11182
Int_rs79494067GA	A	nd	A		nd

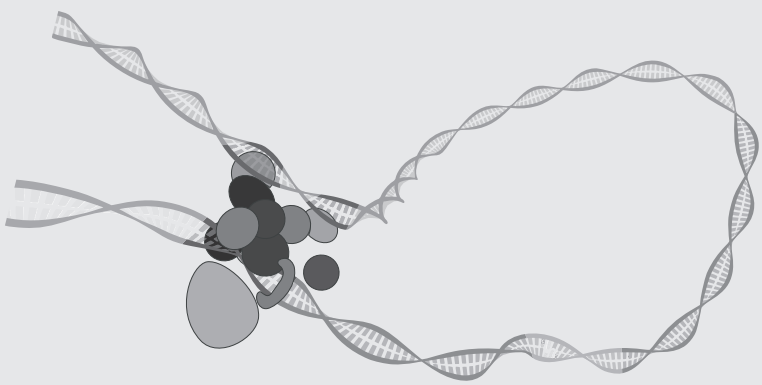
Supplemental Table S2 is available at the Genome Research website.

Chapter 3

Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby *BNC2* pigmentation gene

Mijke Visser, Robert-Jan Palstra, and Manfred Kayser

Human Molecular Genetics, 23, 5750-5762 (2014)



Abstract

Single nucleotide polymorphisms (SNPs) found to be statistically significant when associated with human diseases and other phenotypes are most often located in non-coding regions of the genome. One example is rs10756819 located in the first intron of the *BNC2* gene previously associated with (saturation of) human skin color. Here we demonstrate that a nearby intergenic SNP (rs12350739) in high linkage disequilibrium with rs10756819 is likely the causal DNA variant for the observed *BNC2* skin color association. The highly-conserved region surrounding rs12350739 functions as an enhancer element regulating *BNC2* transcription in human melanocytes, while the activity of this enhancer element depends on the allelic status of rs12350739. When the rs12350739-AA allele is present, the chromatin at the region surrounding rs12350739 is inaccessible and the enhancer element is only slightly active, resulting in low expression of *BNC2,* corresponding with light skin pigmentation. When the rs12350739-GG allele is present however, the chromatin at the region surrounding rs12350739 is more accessible and the enhancer is active, resulting in a higher expression of *BNC2,* corresponding with dark skin pigmentation. Overall, we demonstrate the identification of the functional DNA variant that explains the *BNC2* skin color association signal, providing another important step towards further understanding human pigmentation genetics beyond statistical association. We thus deliver a clear example of how an intergenic non-coding DNA variant modulates the regulatory potential of the enhancer element it is located within, which in turn results in allele-dependent differential gene expression affecting variation in common human traits.

Introduction

Pigmentation of human skin is highly diverse between and within human populations; it varies in color, due to differences in type, amount, and distribution of melanin pigment present in skin melanocytes, and it varies in skin type determined by its UV-sensitivity (1). Many genes have been assigned to this highly complex and polygenic human trait (2). With the latest genetic approaches, such as Genome-Wide Association Studies (GWASs), many single nucleotide polymorphisms from several genes were identified to be statistically significant when associated with variation in human skin pigmentation (3-7). Among the most well-known and confirmed pigmentation genes are *OCA2*, *HERC2*, *ASIP*, *IRF4*, *MC1R*, *TYR*, *TYRP1*, *SLC45A2*, and *SLC24A5* (8). Recently, a candidate-gene approach study identified several SNPs that are suggested to be involved in skin pigmentation pathways, most notably rs10756819 located in the *BNC2* gene (9). Another SNP in this gene (rs2153271) was previously found to be associated with freckling (10), and more recently *BNC2* was also identified to be associated with skin pigmentation in East Asian populations (7). Human skin color is widely assumed to have evolved in humans via adaptation to sunlight as a result of the Out-of-Africa migration of modern humans (11, 12). Indeed, signatures of positive selection were successfully identified in several human pigmentation genes (13-15), supporting this hypothesis. The *BNC2* gene was most recently highlighted as one of the genes present in regions of the human genome that show increased levels of Neanderthal ancestry (16, 17), suggesting that Neanderthals provided modern humans with adaptive variation for skin

phenotypes involving *BNC2* (17).

Basonuclin 2 (BNC2) is one of the most evolutionary-conserved DNA-binding zinc-finger proteins expressed in many human tissues, including epithelial and germ cells. BNC2 consists of three separated pairs of zinc fingers, a nuclear localization signal, and a serine-rich region. Due to its very high conservation status, the function of *BNC2* is suggested to be essential. *BNC2* is most likely involved in mRNA splicing or other forms of mRNA processing (18, 19), but it has also been suggested to function as a transcription factor (20). *BNC2* has 6 promoters and at least 23 (alternative) exons, therefore theoretically resulting in a large number of almost 90,000 possible isoforms, encoding for over 2,000 different proteins. The main human *BNC2* isoform is predicted to encode a 1,099 residue protein with a molecular mass of 122 kDa, but depending on the isoform this can vary tremendously (21). *BNC2* is detected in all layers of the human epidermis, where it resides in nuclear speckles (22), while its paralog *BNC1* is uniformly present in nuclei and was found to be confined to the basal cells of stratified squamous epithelia (22).

BNC2 overexpression studies in mice were the first to implicate this gene in pigmentation. Mice mutants overexpressing *bnc2* were characterized by melanocytic cell death, and subsequent loss of pigment at the base of the hair (23). In addition, *bnc2* knockout mice were shown to be lethal at the embryonic stage or within 24 hours of birth, so that it has not yet been possible to study the role of *bnc2* in mouse pigmentation (23, 24). In zebrafish, the Bonaparte *bnc2* mutant has a normal pigment pattern in adolescence, but the adult stripe pattern is severely disturbed due to extensive loss of pigment cells. Genetic analyses in zebrafish demonstrated that *bnc2* acts non-autonomously to the melanophore lineage and that it is expressed in the hypodermal cells adjacent to pigment cells during adult pattern formation. The Bonaparte *bnc2* mutants are truncated either just before the coding region of the first pair of zinc fingers, leading to the loss of all three pairs of zinc fingers, the NLS and the serine stripe, or it is truncated just after the first pair of zinc fingers, which leaves the protein with only one functional element, resulting in a milder pigment pattern phenotype (25). *Bnc2* is suggested to promote the development and survival of pigment cells in zebrafish by ensuring adequate expression of pigmentation genes like *kitlg* and *csf1* (26).

The exact role of *BNC2,* and in particular that of the pigmentation-associated *BNC2* SNP rs10756819 in human skin pigmentation, has not yet been elucidated. Recently, three studies investigated the functional potential of pigmentation-associated SNPs in three other genomic regions, which led to the identification of enhancer elements regulating transcription of pigmentation genes with a pivotal role for the SNPs located in either a neighboring gene i.e., *HERC2* rs12913832 regulating *OCA2* expression (27) , in an intron of the regulated gene itself i.e., *IRF4* rs12205392 regulating *IRF4* expression (28), or intergenic i.e., rs12821256 regulating *KITLG* expression (29). In the present study, we used several molecular approaches in human skin epidermal tissue and cultured melanocytes to study the pigmentation status of *BNC2*, and the involvement of the associated SNP and its LD partners in the transcriptional regulation of *BNC2* in particular. In multiple melanocyte cell lines derived from differently pigmented skin, as well as in epidermal skin samples from donors of different skin color phenotypes, we studied the expression of the *BNC2* gene and its correlation with their pigmentation phenotypes and genotypes. We profiled the chromatin structure of the *BNC2* region to identify potential regulatory elements using chromatin immune-precip-

itation (ChIP) of H3K27Ac (an active-enhancer mark) in combination with next generation sequencing (ChIP-seq) and ENCODE data. We provide evidence that a region upstream from the canonical *BNC2* promoter acts as an enhancer regulating the expression of *BNC2*, and this transcriptional regulation is dependent on the allelic status of an LD partner of the pigmentation-associated SNP rs10756819. Furthermore, we also provide evidence for the presence of an alternative promoter that stimulates expression specifically of the C-terminal part of *BNC2* that houses the functional elements (e.g. the zinc fingers, the NLS and the serine rich region).

Results

Differential BNC2 expression in skin epidermal samples and in melanocyte cell lines

To study the role of *BNC2* in human pigmentation, and specifically the variation in the transcriptional regulation of this gene, we used a panel of 29 skin epidermal samples obtained from donors with different pigmentation phenotypes (12 dark and 17 light, see supplemental figure S1 for category separation), as well as a melanocytic cell system consisting of 6 commercially available primary Human Epidermal Melanocytes of neonatal origin. The melanocytes were derived from 2 different light pigmented donors (HEMn-LP22 and HEMn-LP89), from 1 moderately pigmented donor (HEMn-MP01) and from 3 darkly pigmented donors (HEMn-DP74, HEMn-DP80 and HEMn-DP83). The characterization of LP22 and DP74 was described previously (27). In addition, HIrisPlex analysis (30) (supplemental tables S1 and S2) of DNA samples from the melanocytic cell lines as well as from the skin epidermal samples, and STRUCTURE analysis (31) (supplemental figures S2) using 24 ancestry-informative autosomal SNPs together with a worldwide reference dataset (HGDP-CEPH) (32), of DNA samples from the melanocytic cell lines confirmed the phenotypic information of the 6 melanocyte-cell donors provided by Cascade Biologics, as well as the phenotypic information provided by the donors of the skin biopsies via a questionnaire.

Our first aim was to ascertain the involvement of *BNC2* in human skin pigmentation by investigating the transcriptional status of *BNC2* in relation to the pigmentation phenotype of the skin epidermal samples. RNA and DNA were co-extracted from the epidermal layers, and mRNA levels of *BNC2* in the individual skin epidermal samples were measured using quantitative (q) RT-PCR. We observed higher expression levels of *BNC2* in the dark skin-colored sample group as compared to the *BNC2* expression levels in the light skin-colored sample group (fig 1A, $p<0.002$). However, the basal layer of the human epidermis consists of keratinocytes and melanocytes, in a ratio of 5:1, overall the keratinocyte percentage in the epidermal layer is even higher (about 95%). Every melanocyte in the basal layer is associated with approximately 36 keratinocytes in the epidermis using their dendrites to transport pigment-containing melanosomes to the keratinocytes (33). Previously it was shown that in zebrafish *bnc2* is expressed in hypodermal cells adjacent to the pigment cells, thereby providing an important environment for pigment cell survival (25, 26). Extrapolating from these results, it was recently suggested that *BNC2* is involved in freckle formation in humans by signaling from keratinocytes to melanocytes (34). In order to investigate whether the

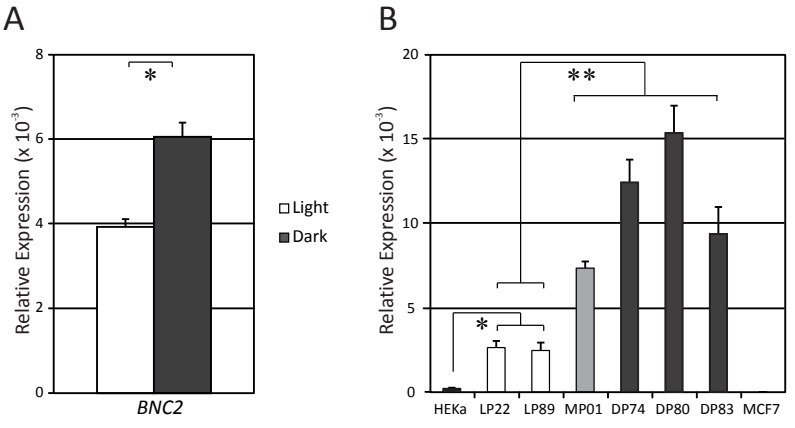


Figure 1. Expression analysis of *BNC2* in skin epidermal samples and in melanocyte cell lines. (A) RT-qPCR analysis of *BNC2* transcripts in light (n=17) and dark (n=12) skin epidermal samples demonstrates differential *BNC2* expression between the light and dark skin-colored sample groups. (B) RT-qPCR analysis of *BNC2* transcripts in 6 melanocyte cell lines, in keratinocytic RNA (HEKa) and in control cell line MCF7. *BNC2* expression is differential between the light (LP22 and LP89) and the moderately and dark (MP01, DP74, DP80 and DP83) melanocyte cell lines. *BNC2* expression is at 10-fold to 70-fold lower in HEKa. Each individual gene expression analysis is carried out in triplicate and normalized to an endogenous control (*ACTB*). Averaged expression values for the two skin color phenotype groups (light and dark) of the skin epidermal samples are calculated after normalization with *ACTB*. Data are represented as mean ± SEM; (*) $p < 0.002$; (**) $P < 0.00004$.

observed differential *BNC2* expression signal in the skin epidermal sample set originates from either the melanocytic or the keratinocytic content, we measured the mRNA levels of *BNC2* in our melanocytic cell line system and, in addition, in human keratinocytic RNA (HEKa, obtained from ScienCell, cat# 2115). This revealed that in humans *BNC2* is expressed at least 10 times higher in melanocytes than in keratinocytes (fig 1B, $p<0.002$). Furthermore, the expression of *BNC2* is higher in the dark and moderately pigmented cells, as compared to the expression of *BNC2* in the light pigmented cells (fig 1B, $p<0.00004$). Expression levels of *BNC2* were very low in MCF7 control cells as may be expected (fig 1B).

From these results we conclude that the *BNC2* gene is expressed in melanocytes, it is differentially expressed between light and dark skin epidermal samples, and this corresponds with the finding that *BNC2* expression levels are higher in the dark pigmented melanocytic cell lines than in the light pigmented melanocytic cell lines. Furthermore, expression of *BNC2* is at least 10-fold higher in human skin melanocytes as compared to human skin keratinocytes, suggesting that *BNC2* has a role in skin pigmentation from within the melanocytic unit of the human epidermis, which is cell autonomous for the melanocytes as opposed to what was suggested previously for zebrafish in which *bnc2* acts on the pigment cells from the surrounding hypodermal cells (25, 26). We therefore consider our melanocytic system a proper system for the investigation of the variation in transcriptional regulation of *BNC2*.

Allele-specific enhancer element upstream of the BNC2 promoter

Jacobs et al. (9) recently identified a pigmentation-associated SNP rs10756819 in the *BNC2* gene. This SNP is located in the first intron of *BNC2* and is therefore not coding for an amino acid. Notably, the region surrounding rs10756819 is highly conserved suggesting functionality. High conservation of non-coding regions might be indicative of regulatory regions and previous studies have demonstrated a role for pigmentation-associated non-coding SNPs in modulating the activity of regulatory regions (27-29). Therefore we reasoned that the position of rs10756819 might coincide with a regulatory element that controls *BNC2* expression levels. Regulatory elements in the genome are characterized by several hallmarks (reviewed recently by Shlyueva et al (35)).

To identify potential enhancer elements that regulate transcription of *BNC2*, we profiled the chromatin of the *BNC2* region considering several data sets that represent features associated with regulatory regions: ChIP-seq analysis in the lightly pigmented melanocytic cell line LP22 and the darkly pigmented melanocytic cell line DP74 of acetylated histone H3 (H3K27Ac) (Palstra et al, manuscript in preparation), an active chromatin mark (36); DNaseI hypersensitive sites in epidermal skin melanocytes (37); ChIP-seq data for the transcription factor MITF in melanocytic cells (38), MITF is the melanocyte master regulator (39); ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1 (40), a ubiquitously expressed transcription factor that was reported to play an important role in melanocyte development by interacting with the melanocyte-specific isoform of MITF (40); predicted melanocyte-specific enhancers (41) and Phastcons conserved elements inferred from 46 way alignments of placental mammals (42). An overview of all tracks in the entire *BNC2* region is shown in supplemental figure S3. The profile reveals several regions that contain potential enhancer features, both in and outside of the *BNC2* gene. However none of these regions with regulatory potential coincide with rs10756819 (figure 2B).

As rs10756819 was identified as a pigmentation-associated SNP by a candidate-gene approach that only used SNPs located within genes, and not in intergenic regions, we reasoned that rs10756819 might be in linkage disequilibrium with the actual causal DNA variant, which was not previously found due to the applied study design (9). We therefore investigated the SNPs that are in high ($R^2 > 0.6$) linkage disequilibrium with rs10756819 (figure S4a, LD plot, and LD SNPs are listed in table S3a); these so-called LD SNPs are located within a narrow frame of 70 kb overlapping the 5' part of *BNC2* (figure 2A). Notably, this LD region is overlapping the LD region of the freckling associated SNP rs2153271 (10). Most of the LD SNPs are not associated with any chromatin marks indicative of regulatory potential, although rs2153271 and rs10810650 flank a region of high H3K27Ac enrichment and DNaseI hypersensitivity corresponding to the *BNC2* promoter region (figure 2A). Interestingly, two LD SNPs, rs10810657 and rs12350739, are located in a region 14 kb upstream of the first and canonical *BNC2* promoter (see figure 2C), which is characterized by the presence of H3K27Ac peaks that are differential between LP22 and DP74, a melanocyte DNaseI hypersensitive signal and a strong conservation signal. Furthermore, an enhancer element was predicted within this region (41) (figure 2C), indicating that this region might be an enhancer with one or two pigmentation-associated SNPs included.

The rs12350739-A derived allele is most common in Europeans (57%), very rare

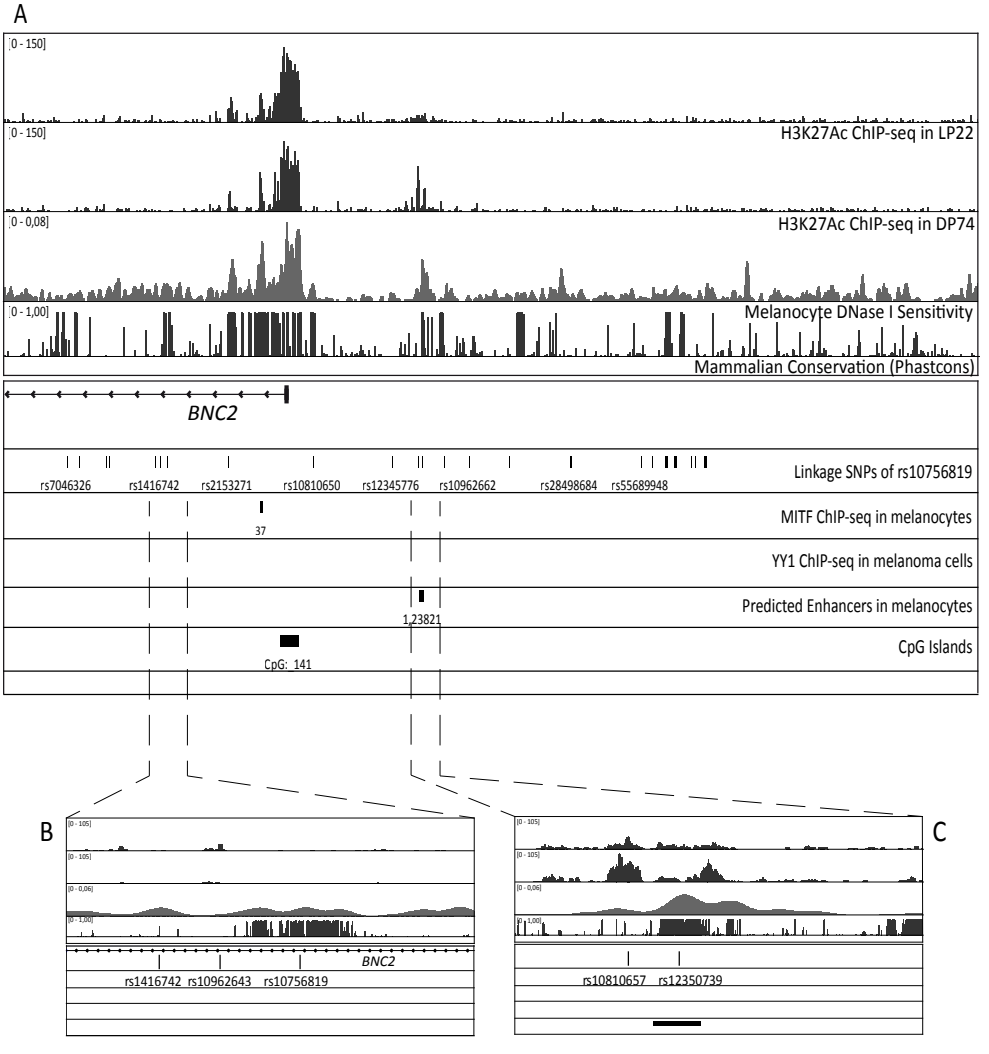


Figure 2. *(legend at the bottom of the next page)*

in Sub-Saharan Africans (1%), and absent in East Asians is evident from the 1000 Genomes project data (43). The rs10810657-A derived allele, however, is found in all populations; its frequency is highest in Europeans (60%) and lowest in sub-Saharan Africans (13%) (43). Remarkably, rs12350739 is classified as a "regulatory region variant", whereas rs10810657 is classified as an "intergenic variant" by the Sequence Ontology project (44, 45). To study the functional potential of these two identified LD SNPs, we investigated the correlation between their genotypes and the *BNC2* expression patterns of the skin epidermal sample set. DNA samples were used for genotyping of the 3 SNPs (supplementary table S4), while expression levels of *BNC2* were measured in the skin epidermis samples by quantitative (q) RT-PCR at exon 5-6. We found a strong correlation between the expression levels of

BNC2 and the genotypes of rs10810657 and rs12350739 (figure 3B-C, $p<0.05$). We also tested the pigmentation-associated SNP rs10756819 and found that the expression levels of *BNC2* did not significantly correlate with the genotypes of this SNP (figure 3A), supporting the observation that rs10756819 does not coincide with a region that contains regulatory potential for *BNC2*, and is therefore most likely not the causal variant for the variation in *BNC2* transcription. Rs10810657 and rs12350739 are in high LD with each other ($R^2 > 0.8$, supplemental figures S4B and S4C, and supplemental tables S3B and S3C); therefore, combinatorial analysis of the rs10810657 and rs12350739 genotypes with the *BNC2*-expression levels resulted in a slightly higher significant correlation between the genotypes and the *BNC2* expression levels (figure 3D, $p<0.01$).

Taken together, these data point towards the presence of a potential regulatory element 17kb upstream of *BNC2*. This region includes two SNPs (rs10810657 and rs12350739) that are in high linkage disequilibrium with the pigmentation-associated SNP rs10756819. The expression levels of *BNC2* correlate strongly with the genotypes of rs10810657 and rs12350739, suggesting that the region around these SNPs potentially acts as an enhancer element.

Characterization of the intergenic BNC2 enhancer rs12350739

Regulatory elements are characterized by open chromatin (46). If the areas around the SNPs in this particular -17kb region include enhancer elements, we should be able to detect accessible chromatin at these locations. We therefore used a commercially available assay that directly measures the accessibility of chromatin to nucleases (EpiQ-assay; BIO-RAD). This assay revealed (moderately) open chromatin in the area around rs12350739 in cell lines harboring the GG or GA genotype, whereas in the cell line with the AA genotype the chromatin was fully closed (figure 4A). Remarkably, the area around rs10810657 was completely inaccessible in all cell lines tested (figure 4B). Enhancer elements characterized by a DNase I hypersensitive peak are often not overlapping with H3K27Ac peaks, but instead are flanked by two prominent H3K27Ac peaks and overlap a less prominent H3K27Ac signal (47). The

Figure 2. Identification of a potential allele-specific enhancer element upstream of the *BNC2* promoter. (A) IGV genome browser (64) shows 100 kb window of the LD region containing the SNPs in linkage disequilibrium (R2 > 0.6) with the pigmentation-associated SNP rs10756819 (hg coordinates chr9:16,840,877-16,943,644). To investigate the chromatin for features of enhancer elements, the following tracks are included: ChIP-seq analysis in LP22 and DP74 of acetylated histone H3 (H3K27Ac), an active chromatin mark (Palstra et al, manuscript in preparation); DNaseI hypersensitive sites in epidermal skin melanocytes (37); ChIP-seq data for the transcription factor MITF in melanocytic cells (38); ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1 (40); predicted melanocyte-specific enhancers (41) and Phastcons conserved elements inferred from 46 way alignments of placental mammals (42). IGV genome browser shows (B) a zoomed-in frame of the region around the pigmentation-associated SNP rs10756819. No enhancer features can be detected in this region. In (C) a zoomed-in frame of the region around potential enhancer element the signals for H3K27Ac are differential between LP22 and DP74, a DNase I peak is detected and an melanocyte-specific enhancer element is predicted here. Two SNPs are located in this region: rs10810657 and rs12350739.

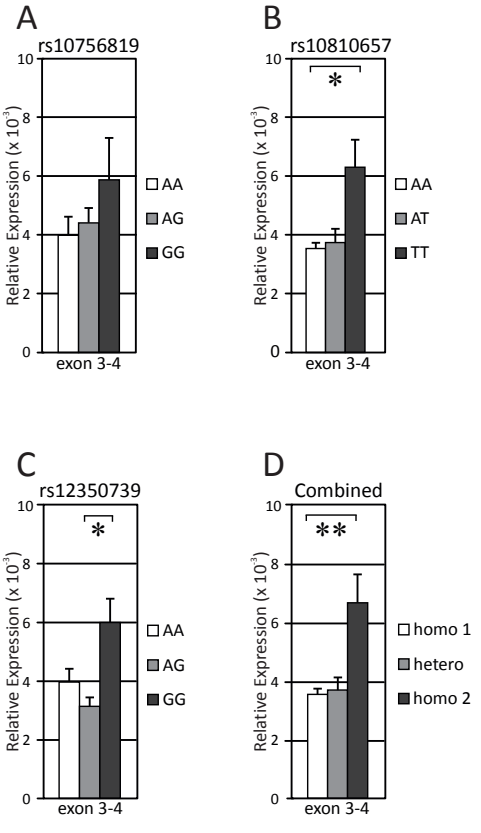


Figure 3. *BNC2* expression depends on the allelic status of rs10810657 and rs12350739. Correlation between the expression patterns of *BNC2* in skin epidermal samples with the genotypes of (A) rs10756819 (n=10,13,6 for AA,AG,GG); (B) rs10810657 (n=8,11,10 for AA,AT,TT); (C) rs12350739 (n=9,8,12 for AA,AG,GG) and (D) rs10810657 and rs12350739 combined (n=8,12,9 for homo 1(AA&AA), hetero (AT&/orAG), homo 2 (TT&GG)). Each individual gene expression analysis is carried out in triplicate and normalized to an endogenous control (*ACTB*). Averaged expression values for the genotype groups are calculated after normalization with *ACTB*. Data are represented as mean ± SEM; (*) $p < 0.05$; (**) $P < 0.01$.

SNP rs10810657 is positioned within one of the more prominent flanking H3K27Ac peaks and at this location the chromatin is not open (DNaseI track in figure 2 and EpiQ results in figure 4B), while the position of rs12350379 is in the middle of the most prominent DNaseI peak and within the predicted melanocyte-specific enhancer element. These observations together with the allelic distribution and the Sequence Ontology project classification suggests that rs10810657 is probably not the regulatory DNA variant involved in transcriptional regulation of *BNC2*, and we therefore focused our further investigations on rs12350379 and its surrounding region.

H3K27Ac ChIP-seq analysis in LP22 and DP74 revealed several H3K27Ac peaks within the *BNC2* region, including the ones in the area around rs12350739 (figure 2C). We confirmed the obtained H3K27Ac ChIP-seq results by ChIP-qPCR in LP22 and in 3 additional cell lines; LP89, DP80 and DP83 (figure 4C). Occupancy of H3K27Ac at rs12350739 was low in both LP22 and LP89, and relatively high in DP80 and DP83 (figure 4C), H3K27Ac occupancy at rs12350739 correspond with the genotypes, the epiQ data, and expression levels of *BNC2* in these cell lines. There was one exception; LP89 had the active and open rs12350739 genotype and epiQ profile, but expressed *BNC2* as low as LP22 did and had a similarly low

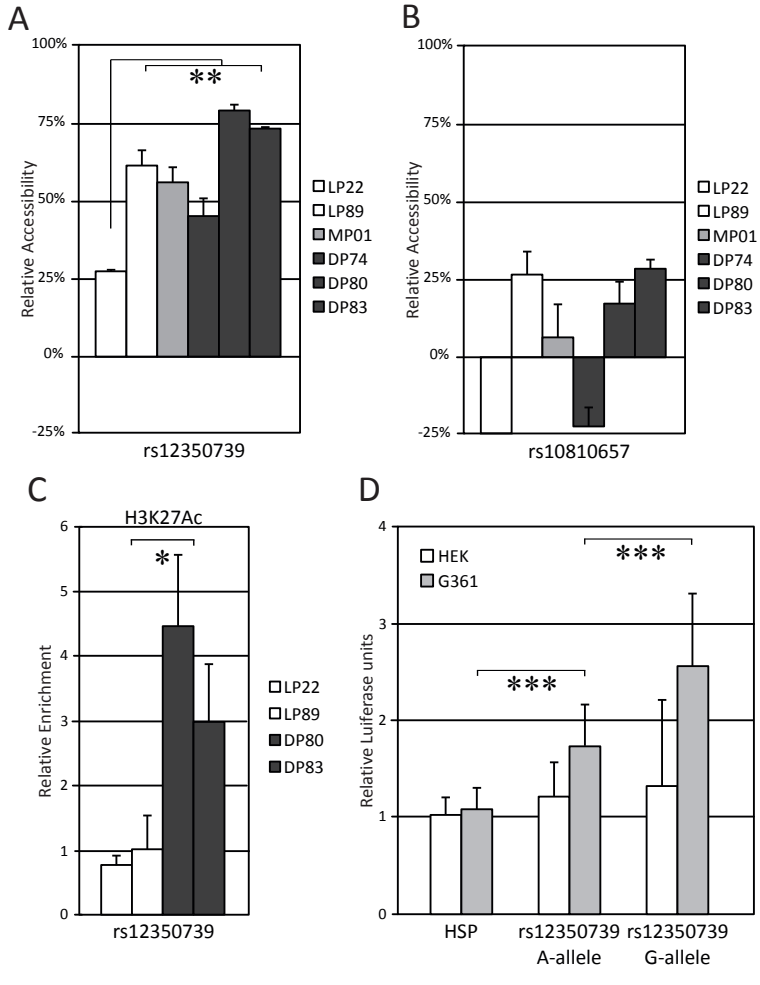


Figure 4. The region around rs12350739 acts as an allele-dependent enhancer. Results of the EpiQ analyses at (A) the region around rs12350739 demonstrating that the chromatin is accessible for LP89 (GG), MP01 (AG), DP74 (AG), DP80 (GG) and DP83 (AG), while the chromatin is inaccessible for LP22 (AA) (** $p < 0.02$) and (B) at the region around rs10810657 demonstrating that the chromatin is fully closed in all cell lines at this SNP. (C) ChIP-qPCR of H3K27Ac in the rs12350739 region demonstrates the potential of an active enhancer present at this region in the dark-pigmented cell lines DP80 and DP83, but not in the light-pigmented cell lines LP22 and LP89 (* $p < 0.05$). (D) Luciferase reporter assay demonstrates that the enhancer activity for the rs12350739 region is differential between the rs12350739-AA allele and the rs12350739-GG allele (*** $p < 0.0005$). Data are represented as mean ± SEM.

H3K27Ac signal as LP22 (see discussion).

To investigate whether the region around rs12350739 acts as an enhancer element we cloned the rs12350739 region into a luciferase reporter vector. Upon transfection into human embryonic kidney cells (HEK293), no substantial increase in luciferase expression

was detected for the vectors containing the rs12350739 region with the different alleles of rs12350739, as compared to the empty vector (figure 4D). However, transfection of the constructs in melanoma G361 cell lines resulted in induced luciferase expression for both alleles, with an 1.5 times higher activity for the rs12350739-GG allele than for the rs12350739-AA allele (figure 4D, $p<0.0005$).

Taken together, these results demonstrate that the features of the rs12350739 region are consistent with those of an enhancer element, and the activity of this region depends on the allelic status of rs12350739.

Additional regulatory elements within the BNC2 locus and an alternative BNC2 promoter

In the initial H3K27Ac chromatin profile of the *BNC2* locus, we observed several H3K27Ac peaks that could represent additional regions with regulatory potential. We selected 4 regions (marked A-D in supplemental figure S3) for further analysis based on the following criteria; we included regions if they contained a robust H3K27Ac peak and a DNaseI hypersensitivity signal and if they displayed MITF binding. We investigated the potential contribution of these 4 regions (A-D) in determining pigmentation via the regulation of *BNC2* transcription using epiQ analysis, in combination with ChIP-qPCR for the active enhancer mark H3K27Ac. Region A involves the canonical promoter region of *BNC2*; we profiled region A with epiQ; the chromatin of the *BNC2* promoter itself was moderately open in all cell lines, and decreased with distance as such that the chromatin at SNP rs10756819 was closed (supplemental figure S5A). The association of rs10756819 with the expression pattern of *BNC2* in skin epidermal samples was not statistically significant, whereas the association of rs2153271 with the expression pattern of *BNC2* in skin epidermal samples was highly statistically significant ($p<0.001$, data not shown). This was most probably due to the high linkage of rs2153271 with rs12350739 since the region around rs2153271 lacks chromatin features indicating regulatory potential. Two regions with regulatory potential stand out in the 3' region of the *BNC2* gene. Region C is located in intron 5 of *BNC2* and region D is intergenic, approximately 80 kb downstream from *BNC2* (supplemental figure S5B). EpiQ analysis revealed moderately open chromatin in both regions, which did not show correlation with *BNC2* expression.

The most prominent H3K27Ac peak within the *BNC2* locus is represented by region B. This region encompasses a MITF binding site together with a tandem DNaseI hypersensitive peak and two YY1 sites suggesting that this region actually consists of two regulatory elements B' and B". EpiQ analysis revealed that the chromatin in region B' was equally accessible in all cell lines, while accessibility was even more significant in the sub-region B" (figure 5A). Region B' exhibited high and differential H3K27Ac ChIP-qPCR signals among the cell lines tested; LP22 and LP89 displayed low H3K27Ac enrichment as compared to DP80 and DP83 (figure 5B).

Notably, this corresponds with the expression pattern of *BNC2*; expression of *BNC2* is low with low H3K27Ac-enrichment in LP22 and LP89, as compared to higher H3K27Ac-enrichment and higher expression of *BNC2* in DP80 and DP83. Ensembl gene prediction suggests that several protein-coding transcripts originate in this region (ENST00000418777;

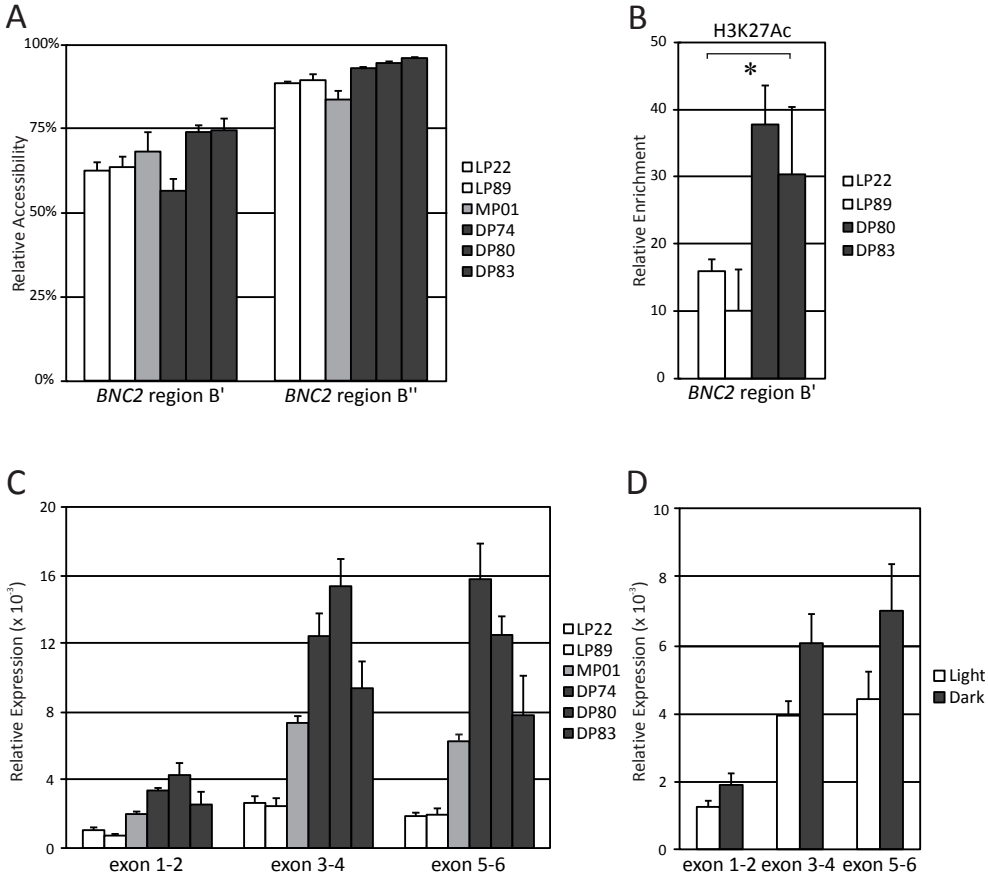


Figure 5. The identification of an alternative promoter in the *BNC2* gene. (A) Results of the EpiQ analysis demonstrates that the chromatin in regions B' and B'' is highly accessible. (B) The H3K-27Ac-ChIP-qPCR signals in region B' are high and differential between the light-pigmented (LP22 and LP89) and dark-pigmented (DP80 and DP83) cell lines (* $p < 0.05$). (C) RT-qPCR analysis of *BNC2* transcripts in 6 melanocyte cell lines at one location upstream and 2 locations downstream of the alternative promoter within the *BNC2* locus demonstrates that *BNC2* expression is higher downstream of the alternative promoter (* $p < 0.05$), confirming the RNA-seq data. (D) RT-qPCR analysis of *BNC2* transcripts in light (n=17) and dark (n=12) skin epidermal samples demonstrates differential *BNC2* expression between the upstream location and the two downstream locations within the *BNC2* gene (** $p < 0.0005$). Data are represented as mean ± SEM.

ENST00000603713; ENST00000603313 and ENST00000468187) (45). Therefore, we suggest that this region potentially acts as an alternative *BNC2* promoter, rather than an enhancer element. To test this hypothesis, we investigated the occupancy of the major promoter-associated chromatin mark H3K4Me3 in the *BNC2* region in two publicly available H3K4Me3 ChIP-seq datasets obtained from human skin melanocytes . This revealed two strong peaks in the *BNC2* region, one at the canonical promoter and one at region B'. We confirmed the occupancy of H3K4Me3 by ChIP-qPCR in the light-pigmented cell line LP89 at the canonical

promoter (supplemental figure S6A) and at region B' in *BNC2* (supplemental figure S6B).

Binding of RNA Polymerase II, similarly tested with ChIP-qPCR in LP89, is relatively high at the alternative promoter and region B", which has several features consistent with an enhancer element, e.g. high H3K27Ac enrichment, low H3K4Me3 enrichment and binding of MITF and YY1 (supplemental figure S6B). At the canonical promoter binding of RNA polymerase II is relatively low (supplemental figure S6A), suggesting that transcription is higher downstream of the alternative promoter. We therefore analyzed RNA-seq data we obtained from 4 melanocyte cell lines (LP89, DP80, DP74 and MP01). This revealed that the majority of transcripts initiated at exon 3 and not at exon 1 (supplemental figure S6). We confirmed this finding by RT-qPCR analysis of *BNC2* expression levels at 3 different locations within the gene, using one primer set targeting mRNA upstream of the alternative promoter, and 2 primer sets targeting the mRNA downstream of the alternative promoter. Transcript levels measured upstream from the alternative promoter were significantly lower in all cell line samples, as compared to transcript levels downstream from the alternative promoter measured at two different locations, while expression patterns remained similar among the primer sets for all cell lines tested (see figure 5C, $p<0.02$). Similar results were obtained when testing these three primer sets in the skin epidermal samples (figure 5D, $p<0.0004$).

In addition, we have noted the presence of a CpG island located within the potential alternative promoter region (see supplemental figure S6). Unless under positive selection, methylated C nucleotides tend to be eliminated during evolution and hence CpG islands are frequently found in TATA-less promoters (48), as is the case for the canonical promoter of *BNC2*. Therefore the presence of this CpG island in exon 3 of *BNC2* supports our notion that region B' acts as an alternative *BNC2* promoter.

From these findings we conclude that an alternative promoter is present at exon 3 in *BNC2*, as the majority of *BNC2* transcripts initiate from this location, and not from the canonical promoter. Furthermore, the chromatin features are consistent with active enhancers and/or promoters, transcription factor binding sites for MITF and YY1 are present and a CpG island is located in this region.

Discussion

GWASs provide extensive lists of SNPs statistically significantly associated with a variety of human traits including pigmentation. Many of these highlighted associated DNA variants are located within non-coding regions (intronic or intergenic) of the genome, and are therefore not directly affecting protein functions by altering the amino-acid sequences. Therefore, the biological function behind such statistical associations often remains unknown since the association can stem from impeding another biological, e.g. regulatory function, or from linkage to another (functional) SNP (49, 50) probably due to the design of the SNP array used, and this can only by elucidated by subsequent experimental studies. Here we studied the biology behind a recently identified skin color association signal in *BNC2* and found that it was not the associated SNP rs10756819 itself but its LD partner rs12350739 located 14 kb upstream of *BNC2* that was likely the functional DNA variant responsible for the skin color association signal of *BNC2*.

We demonstrated that the highly conserved region surrounding rs12350739 func-

tions as an enhancer element for the nearby *BNC2* gene, and that this function is dependent on the allelic status of rs12350739. When the rs12350739-AA allele is present, the chromatin at the region surrounding rs12350739 is fully inaccessible and the enhancer element was only moderately active. As a consequence, skin epidermal samples with the AA-allele display a low expression of the *BNC2* gene, which corresponds with a light skin pigmentation phenotype. When the rs12350739-GG allele is present however, the chromatin at the enhancer is more accessible and the enhancer is active. This results in a higher expression of *BNC2*, corresponding with a dark skin pigmentation phenotype.

We have identified rs12350739 as the causal i.e., regulatory DNA variant for the *BNC2* skin color effect by studying the linkage partners of rs10756819, the particular *BNC2* SNP that was previously found to be highly associated with skin pigmentation (9). It can be argued however, that other linkage SNPs are contributing to the variation in transcriptional regulation of *BNC2*. We have investigated the region containing the LD block in detail, considering the hallmarks of active enhancers, and found that only three SNPs are located in the vicinity of potentially regulatory regions; rs12350739, rs10810657 and rs2153271. Of these three SNPs, only rs12350739 turned out to be regulatory, and we consider it highly unlikely that other SNPs are functionally contributing to the transcriptional regulation of *BNC2* in skin melanocytes.

Several candidate regulatory elements are present within the *BNC2* gene, and it might be suggested that one of these elements is responsible for the regulation of *BNC2* transcriptional variation. However, we find this hypothesis unlikely for several reasons. First, *BNC2* was shown to be a pigmentation gene (10, 23, 25), and the SNP associated with skin pigmentation and the LD block linked to rs10765819 are located within a narrow frame of 70kb surrounding the canonical promoter of *BNC2*. In this window, only two regions displayed features of enhancer elements; the promoter and the identified enhancer, and we clearly demonstrated that the activity of this enhancer depends on the allelic status of rs12350739, which in turn corresponds with the expression levels of *BNC2*. Second, we also showed that the other candidate regions might display features of enhancers, but these features are not linked to the expression patterns of *BNC2*. However, we cannot fully exclude the possibility that other enhancer elements contribute to the variation in transcription of *BNC2*, although in skin melanocytes this potential contribution is probably minor compared to the effect of the rs12350739 enhancer element we identified here.

Previously, it had been suggested that *BNC2* acts on pigmentation through signaling from keratinocytes, rather than from melanocytes (10, 25, 34). We demonstrated here that *BNC2* is differentially expressed in melanocyte cell lines originating from differently pigmented donors, which suggests that *BNC2* does act on pigmentation through melanocytes, at least in humans. We showed that the expression levels of *BNC2* in melanocytes depend on the allelic status of *BNC2* rs12350739, which in turn corresponds with the pigmentation phenotype. As was mentioned in the results chapter, there is one exception in our data; the light-pigmented cell line LP89 expresses *BNC2* at low levels but carries the dark skin-color associated rs12350739-GG allele. Interestingly, genotype-based HIrisPlex analysis predicted that the LP89-donor has dark eyes and dark hair, even though LP89 is classified by the supplier as a light-pigmented cell line. Moreover, the expression levels of several pigmentation genes is reduced in LP89 as compared to those in the dark pigmented cell lines (data not

shown). EpiQ data demonstrates that chromatin at the rs12350739 enhancer region in LP89 is remodeled which is in agreement with the rs12350739-GG genotype. However, in LP89 acetylation of nucleosome H3 at lysine 27 at this region is compromised resulting into similarly low levels as in LP22 (AA-allele), which is in agreement with the *BNC2* expression data. These findings for LP89 suggest that besides the identified rs12350739 enhancer, other, presumably *trans*-acting factors contribute to the transcriptional regulation of *BNC2*, acting at a step between chromatin remodeling and nucleosome acetylation. However, more work is needed to fully elucidate this mechanism.

BNC2 is suggested to function in mRNA processing (18, 20) and it was proposed that in zebrafish it targets the pigmentation genes *kitlg* and *csf1* (26). Although unraveling the biological function of *BNC2* was not the scope of this study, we can confirm a positive correlation between the expression levels of *BNC2* and the 2 proposed targets *KITLG* and *CSF1* using our RNA-seq data in a subset of 5 human melanocyte cell lines (LP74, LP89, MP01, DP74 and DP80, data not shown). This correlation was absent for several other pigmentation genes.

BNC2 is known to have 6 promoters and many splicing isoforms, resulting in a potential of nearly 90,000 mRNA isoforms encoding more than 2000 different proteins (21). Our data clearly shows the presence of an alternative promoter initiating transcription at exon 3, confirming previous data that suggest the presence of this promoter at the same location together with the presence of 5 other minor human promoters for the *BNC2* gene (15). DNaseI sensitivity data in many different tissues suggest that this alternative promoter is very abundant (37, 51), but interestingly, the DNaseI sensitivity peak 6 kb upstream of the alternative promoter that we observed in human skin melanocytes (supplemental figure S3) seems to be tissue-specific as it is only detected in melanoma cell lines (37, 51). It co-localized with H3K27Ac signals in our melanocyte cells and with binding sites for the transcription factors MITF and YY1, both essential for melanocyte biology. Furthermore, H3K4Me3 occupancy is low, making it unlikely that region B" acts as another promoter. Instead, RNA polymerase II binding at this region is similar to that at the alternative promoter and to that at the rs12350739 enhancer element. It is therefore likely that this region acts as a tissue-specific enhancer element for the alternative promoter. We found a relatively modest, allele-dependent increase in luciferase expression for the rs12350739 enhancer element. It has been reported previously that genes can simultaneously be regulated by multiple enhancers which affects gene-expression amplitude (52). Likewise, the genotype-dependent enhancer at rs12350739 in combination with the potential enhancer at region B" could boost *BNC2* transcription by both acting on the alternative promoter which might lead to a differentially expressed isoform that has a specific function in pigmentation pathways.

The DNaseI sensitivity peak that we observed for the enhancer element at rs12350739 is also present in a few other cell types, including fibroblasts, a glioblastoma cell line, melanoma cells and myometrial cells (37, 51). *BNC2* was shown to be involved in urethral development (53), and recently, an eQTL study identified *BNC2* to have a role in glioblastoma multiforme (54). Furthermore, 12 SNPs in the *BNC2* region were identified to be associated with ovarian cancer susceptibility and with ovarian abnormalities (55, 56). Notably, these 12 SNPs all fall into the same LD block as rs12350739, and with a DNaseI peak present at the rs12350739 enhancer in myometrial cells, we suggest that the enhancer we

identified at the region around rs12350739 in skin melanocytes is also active in uterine or ovarian tissue, which should be followed up in dedicated studies.

Several studies have shown that strongly associated SNPs identified by GWASs can have important regulatory functions. They are often located in enhancer elements, influencing the transcriptional regulation of the nearby gene in an allele-dependent manner (recently reviewed in (57-59)). So far however, only the biological functions of particular SNPs that were identified directly in the GWASs were elucidated. Here we demonstrate with the example of *BNC2* and skin color that it was not the SNP directly identified by association testing, but instead one of its LD partners that is likely the regulatory DNA variant responsible for the association signal. This DNA regulator is located in a highly-conserved enhancer element outside the associated gene and regulates gene transcription depending on its allelic status, and the resulting differential gene expression in turn affects the trait variation.

Materials and Methods

Skin sample collection

Skin samples were obtained with informed consent of left-over material from patients undergoing plastic or dermatological surgery. Care was taken to only collect healthy skin tissue that was non-sun exposed and not tanned by solariums. Patients were asked to fill in a questionnaire to obtain information concerning the pigmentation of their skin, eye and hair, and more basic information as age, sex and geographical ancestry. Skin color reflectance was measured at the inner side of the upper arm with a spectrophotometer (Konica Minolta CM-600d). Individual typology angles (60, 61) were calculated with the obtained L, a and b values, and samples were categorized according to the ITA separation of light (above 41 degrees) and dark (under 28 degrees), for a graphical display of the ITA categories of the skin epidermal sample, see supplemental figure S1. This study was approved by the Medical Ethical Committee (METC) of the Erasmus MC (number MEC-2012-067).

Epidermal layer separation and RNA/DNA co-isolation from skin samples

All obtained small skin biopsies were immediately stored in RNAlater at 4°C until further processing. Upon receiving the larger pieces of surgically-removed skin tissue, all tissue other than skin (blood, fat etc) were removed while working on ice, after which the skin was cut into smaller pieces of 1 cm^2 on average, and stored in RNAlater at 4°C until further processing. The epidermal layer of the skin, including the pigment-containing melanocytic layer, was removed from the dermal layer by first cutting the skin into pieces of around 1mm wide and 7mm long and incubating the skin pieces for 30 minutes at room temperature in Ammonium Thio Cyanate solution (3.8% ATC in PBS) (62). After carefully peeling off the epidermal layer from the dermal layer, the epidermal skin layers were stored in RNAlater at 4°C until RNA/DNA co-isolation, or at -80°C for long-term storage. Epidermal skin tissue was lysed and RNA and DNA were co-isolated with the Qiagen Allprep mini kit according to the manufacturer's instructions for fibrous tissue. To remove PCR inhibiting substances such as melanin, DNA and RNA samples were column purified (*OneStep*™ PCR Inhibitor Removal Kit,

Zymo Research Corporation) following the manufacturer's instructions. Subsequent DNase I digestion of the RNA samples was performed with Ambion's Turbo DNA-free kit (Applied Biosystems) according to the manufacturer's protocol.

Cell culture

HEMn-LP22 (C-0025C; lot # 200708522; Cascade Biologics, Invitrogen), HEMn-LP89 (C-0025C; lot # 200709589; Cascade Biologics, Invitrogen), HEMn-MP01 (C-1025C; lot# 070417901; Cascade Biologics, Invitrogen), HEMn-DP74 (C-2025C, lot # 6C0474; Cascade Biologics, Invitrogen), HEMn-DP80 (C-2025C, lot # 200707980 Cascade Biologics, Invitrogen) and HEMn-DP83 (C-2025C, lot # 1211783; Cascade Biologics, Invitrogen) were grown in Medium 254 supplemented with HMGS according to the manufacturer's instruction (Cascade Biologics, Invitrogen). G361, HEK293 and MCF7 cells were cultured in DMEM/10% FCS at 37°C/5% CO2. Transfection was done using Lipofectamine LTX transfection reagent (Invitrogen) according to the manufacturer's instructions.

RNA/DNA co-isolation from cultured cells

Total cellular RNA and genomic DNA were isolated from the different cell lines with TriPure Isolation Reagent according to the manufacturer's instructions (Roche Diagnostics) To remove PCR inhibiting substances such as melanin, DNA and RNA samples were column purified (*OneStep*™ PCR Inhibitor Removal Kit, Zymo Research Corporation) following the manufacturer's instructions. Subsequent DNase I digestion of the RNA samples was performed with Ambion's Turbo DNA-free kit (Applied Biosystems) according to the manufacturer's protocol.

SNP genotyping

To obtain hair and eye color information for the 6 cell line samples and to confirm the obtained eye and hair color phenotype information of all skin biopsy samples, DNA samples of both sets were genotyped using the HIrisPlex assay as described by Walsh et al 2013 (30). The interactive HIrisPlex prediction tool was used to predict hair and eye color (30). Snapshot analysis was used to genotype *BNC2* SNPs. In short, primers were designed using primer3plus software, checked with AutoDimer for unfavorable primer interactions, and, in case necessary, optimized by increasing primer concentrations. Primer sequences are listed in supplemental table S5. Multiplex PCR amplification was performed using GeneAmp PCR Gold Buffer and AmpliTaq Gold DNA polymerase (Applied Biosystems), under the following cycling conditions: 10 min at 95ºC, followed by 30 cycles of 15s at 94 ºC and 45s at 60ºC, and a final extension of 5 min at 60 ºC in a dual 364-well GeneAmp PCR System 9700 (Applied Biosystems). After ExoSAP-IT (USB) purification, multiplex single-base primer extension reactions were performed using SNaPshot Ready Reaction Mix (Applied Biosystems) in a dual 384-well GeneAmp PCR System 9700 (Applied Biosystems) under the following cycling conditions: 2 min at 96 ºC, followed by 25 cycles of 10 s at 96 ºC, 5s at 50 ºC, and 30s at 60 ºC. After SAP-purification of the reaction products, the extended primers were analyzed by capillary electrophoresis on a 3130xl Genetic Analyzer (Applied Biosystems) using POP-7

polymer. Data were analyzed using GeneMarker version 2.4.0 software (Softgenetics). To infer the geographic origin/genetic ancestry of the cell-line donors, 24 autosomal SNPs sensitive to detect continental ancestry were genotyped via two SNaPshot (Applied Biosystems) multiplex reactions as described elsewhere (32). The continental ancestry of the samples was recovered by performing a STRUCTURE analysis (31) in which data from the HGDP-CEPH panel served as reference as described elsewhere (32).

Chromatin Immuno-Precipitation quantitative (ChIP-qPCR) and sequencing (ChIP-seq)

ChIP was performed as described in the Millipore protocol (http://www.millipore.com/userguides/tech1/mcproto407), except that samples were cross-linked with 2% formaldehyde for 10 minutes at room temperature. To remove melanin, DNA was column purified prior to PCR (*OneStep*™ PCR Inhibitor Removal Kit, Zymo Research). Quantitative real-time PCR was performed using Universal SYBR green master mix (Bio-Rad Laboratories) under the following cycling conditions: 95 °C for 5 min, 45 cycles of 10 s at 95 °C, 30 s at 60 °C and followed by a melting curve analysis. Enrichment was calculated relative to Necdin (*NDN*) and values were normalized to input measurements. Antibodies used: acetylated Histone H3 K27 (#4729ab) from Abcam, tri-methylated Histone H3 K4 (#07-473) from Millipore and Polymerase II (POLR2A, N-20; sc-899) from Santa Cruz Biotechnology. Primer sequences are listed in supplemental Table S5. Data represent at least two independent experiments. Student's two-tailed *t* test was used to determine statistical significance. ChIP-seq analysis was performed as described previously (63).

EpiQ

Q-PCR–based EpiQ chromatin analysis assay (Bio-Rad Laboratories) was performed according to the manufacturer's protocol. To remove melanin, DNA was column purified prior to PCR (*OneStep*™ PCR Inhibitor Removal Kit, Zymo Research). Quantitative real-time PCR was performed using Universal SYBR green master mix (Bio-Rad Laboratories) under the following cycling conditions: 95 °C for 5 min, 45 cycles of 10 s at 95 °C, 30 s at 60 °C and followed by a melting curve analysis. PCR primers are listed in supplemental table S5. Data represent at least two independent experiments. Student's two-tailed *t* test was used to determine statistical significance.

Transcription analysis

The reverse-transcriptase (RT) reaction was performed using RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas GmbH) according to the manufacturer's instructions. Quantitative real-time PCR reactions for gene-expression analysis were performed using the iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories) with the following PCR parameters: initial denaturation at 95 °C for 5 min, followed by 45 cycles at 95 °C for 5 s and 60 °C for 30 s, followed by a melting curve analysis. Bio-Rad Software v1.5 in combination with Excel was used to analyze the qPCR data. The reference gene *ACTB* was used to normalize the amplification signal between samples, differences in treatment and amount of input

cDNA. Primer sequences are listed in supplemental Table S5. Data represent at least three independent qPCR experiments. Student's two-tailed *t* test was used to determine statistical significance.

Luciferase Assays

A 1524bp fragment surrounding rs12350739 was PCR amplified from genomic DNA using the Expand long template PCR kit (Roche). The PCR fragment was digested with HaeI and SacI, generating a 599bp fragment which was cloned into the HaeI/SacI sites of a modified pGL3-promoter vector (in which the SV40 promoter and the SV40 3'UTR were replaced by an HSP promoter and an HSP 3'UTR) (Promega). Inserts in each construct were verified by sequencing (Baseclear). Constructs were transfected into HEK293 or G361 melanoma cells using Lipofectamine LTX (Invitrogen), and luciferase expression was normalised to Renilla luciferase expression. Primer sequences are listed in supplemental Table S5. Data represent at least five independent experiments. Student's two-tailed *t* test was used to determine statistical significance.

RNA sequencing

Two hundreds to five hundreds ng of total RNA extracted from the melanocyte cell lines LP89, MP01, DP74 and DP80 was used as starting input for the RNA-seq library preparation, following the protocols provided by Life Technologies. Briefly, total RNA was treated with RiboMinus Eukaryote kit v2 (Life Technologies) to remove rRNA. Then, rRNA-depleted RNAs were fragmented using RNaseIII. Whole transcriptome library was constructed using the Ion Total-RNA Seq Kit v2 (Life Technologies). Each library template was clonally amplified on Ion Sphere Particles (Life Technologies) using Ion One Touch 200 Template Kit v2 (Life Technologies). Preparations containing the RNA-seq libraries were loaded into 318 Chips and sequenced on the PGM (Life Technologies).

Acknowledgments

We are very grateful to all volunteers who agreed their materials being used for this study. We thank the staff, especially Dr Jan Maerten Smit and Dr Han van Neck, from the Department of Plastic Surgery, and Professor Tamar Nijsten, Leonie Jacobs and Emmilia Dowlatshahi from the Department of Dermatology, both of Erasmus MC, who helped us with collecting all skin epidermal samples. We thank the staff of the Erasmus Center of Biomics, especially Wilfred van IJcken and Rutger Brouwer, for sequencing and data-processing of the ChIP-seq samples. Lakshmi Chaitanya and Arwin Ralf from the Department of Forensic Molecular Biology of Erasmus MC are acknowledged for help with genotyping, and Arwin Ralf additionally for help in ancestry inferences, Petros Kolovos from the Department of Cell Biology of Erasmus MC for help with sequencing data analyses and Susan Walsh from the Department of Anthropology, Yale University for valuable comments on the manuscript.
Funding: This study was supported by the Erasmus MC, and a grant from the Netherlands Genomics Initiative (NGI) / Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

References

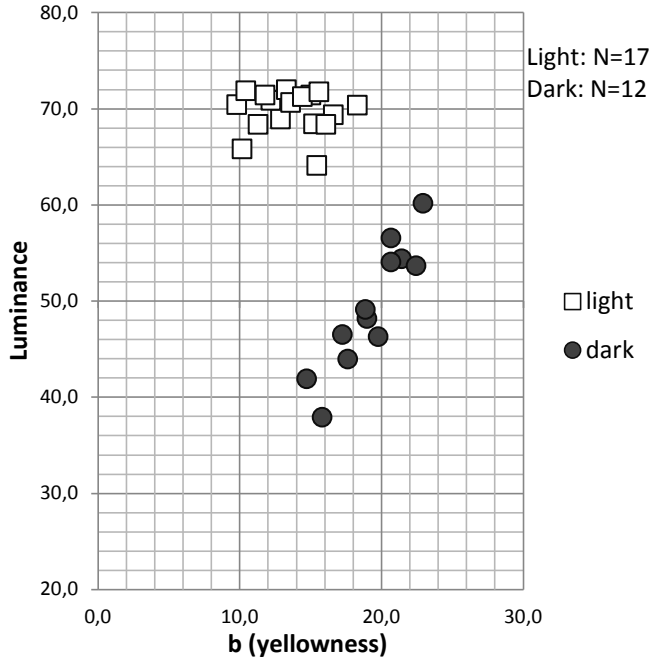
1. Sturm, R.A. (2009) Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.,* **18**, R9-R17.
2. Barsh, G.S. (2003) What controls variation in human skin color? *PLoS Biol.,* **1**, 19-22.
3. Han, J.L., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z. *et al.* (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.,* **4**, e1000074.
4. Nan, H., Kraft, P., Qureshi, A.A., Guo, Q., Chen, C., Hankinson, S.E., Hu, F.B., Thomas, G., Hoover, R.N., Chanock, S. *et al.* (2009) Genome-Wide Association Study of Tanning Phenotype in a Population of European Ancestry. *J. Invest. Dermatol.,* **129**, 2250-2257.
5. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Jakobsdottir, M., Steinberg, S., Gudjonsson, S.A., Palsson, A., Thorleifsson, G. *et al.* (2008) Two newly identified genetic determinants of pigmentation in Europeans. *Nat. Genet.,* **40**, 835-837.
6. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G. *et al.* (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.,* **39**, 1443-1452.
7. Hider, J.L., Gittelman, R.M., Shah, T., Edwards, M., Rosenbloom, A., Akey, J.M. and Parra, E.J. (2013) Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol. Biol.,* **13**, 150.
8. Sturm, R.A. and Duffy, D.L. (2012) Human pigmentation genes under environmental selection. *Genome Biol.,* **13**, 248.
9. Jacobs, L.C., Wollstein, A., Lao, O., Hofman, A., Klaver, C.C., Uitterlinden, A.G., Nijsten, T., Kayser, M. and Liu, F. (2013) Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Hum. Genet.,* **132**, 147-158.
10. Eriksson, N., Macpherson, J.M., Tung, J.Y., Hon, L.S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I. and Mountain, J. (2010) Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. *PLoS Genet.,* **6**, e1000993.
11. Jablonski, N.G. and Chaplin, G. (2000) The evolution of human skin coloration. *J. Hum. Evol.,* **39**, 57-106.
12. Relethford, J.H. (1997) Hemispheric difference in human skin color. *Am. J. Phys. Anthropol.,* **104**, 449-457.
13. Lao, O., de Gruijter, J.M., van Duijn, K., Navarro, A. and Kayser, M. (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann. Hum. Genet.,* **71**, 354-369.
14. Parra, E.J. (2007) Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am. J. Phys. Anthropol.,* **134**, 85-105.
15. Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterlander, M., Hollfelder, N., Potekhina, I.D., Schier, W., Thomas, M.G. et al. (2014) Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. USA*, **111**, 4832-4837.
16. Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N. and Reich, D. (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature,* **507**, 354-357.
17. Vernot, B. and Akey, J.M. (2014) Resurrecting surviving Neandertal lineages from modern human genomes. *Science,* **343**, 1017-1021.
18. Romano, R.A., Li, H.X., Tummala, R., Maul, R. and Sinha, S. (2004) Identification of Basonuclin2, a DNA-binding zinc-finger protein expressed in germ tissues and skin keratinocytes. *Genomics,* **83**, 821-833.
19. Vanhoutteghem, A. and Djian, P. (2004) Basonuclin 2: An extremely conserved homolog of the zinc finger protein basonuclin. *Proc. Natl. Acad. Sci. USA,* **101**, 3468-3473.
20. Herve, F., Vanhoutteghem, A. and Djian, P. (2012) [Basonuclins and DISCO proteins: regulators of development in vertebrates and insects] Basonuclines et proteines DISCO: des regulateurs du developpement des vertebres et des insectes. *Med. Sci. (Paris),* **28**, 55-61.
21. Vanhoutteghem, A. and Djian, P. (2007) The human basonuclin 2 gene has the potential to generate nearly 90,000 mRNA isoforms encoding over 2000 different proteins. *Genomics,* **89**, 44-58.

22. Vanhoutteghem, A. and Djian, P. (2006) Basonuclins 1 and 2, whose genes share a common origin, are proteins with widely different properties and functions. *Proc. Natl. Acad. Sci. USA,* **103**, 12423-12428.
23. Smyth, I.M., Wilming, L., Lee, A.W., Taylor, M.S., Gautier, P., Barlow, K., Wallis, J., Martin, S., Glithero, R., Phillimore, B. *et al.* (2006) Genomic anatomy of the Tyrp1 (brown) deletion complex. *Proc. Natl. Acad. Sci. USA,* **103**, 3704-3709.
24. Vanhoutteghem, A., Maciejewski-Duval, A., Bouche, C., Delhomme, B., Herve, F., Daubigney, F., Soubigou, G., Araki, M., Araki, K., Yamamura, K. *et al.* (2009) Basonuclin 2 has a function in the multiplication of embryonic craniofacial mesenchymal cells and is orthologous to disco proteins. *Proc. Natl. Acad. Sci. USA,* **106**, 14432-14437.
25. Lang, M.R., Patterson, L.B., Gordon, T.N., Johnson, S.L. and Parichy, D.M. (2009) Basonuclin-2 Requirements for Zebrafish Adult Pigment Pattern Development and Female Fertility. *PLoS Genet.,* **5**, e1000744.
26. Patterson, L.B. and Parichy, D.M. (2013) Interactions with Iridophores and the Tissue Environment Required for Patterning Melanophores and Xanthophores during Zebrafish Adult Pigment Stripe Formation. *PLoS Genet.,* **9**, e1003561.
27. Visser, M., Kayser, M. and Palstra, R.J. (2012) HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.,* **22**, 446-455.
28. Praetorius, C., Grill, C., Stacey, S.N., Metcalf, A.M., Gorkin, D.U., Robinson, K.C., Van Otterloo, E., Kim, R.S.Q., Bergsteinsdottir, K., Ogmundsdottir, M.H. *et al.* (2013) A Polymorphism in IRF4 Affects Human Pigmentation through a Tyrosinase-Dependent MITF/TFAP2A Pathway. *Cell,* **155**, 1022-1033.
29. Guenther, C.A., Tasic, B., Luo, L., Bedell, M.A. and Kingsley, D.M. (2014) A molecular basis for classic blond hair color in Europeans. Nat. Genet., in press, doi: 10.1038/ng.2991.
30. Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W. and Kayser, M. (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.,* **7**, 98-115.
31. Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155,** 945-959.
32. Lao, O., Vallone, P.M., Coble, M.D., Diegoli, T.M., van Oven, M., van der Gaag, K.J., Pijpe, J., de Knijff, P. and Kayser, M. (2010) Evaluating Self-declared Ancestry of US Americans With Autosomal, Y-chromosomal and Mitochondrial DNA. *Hum. Mutat.,* **31**, E1875-E1893.
33. Fitzpatrick, T.B. and Breathnach, A.S. (1963) [the Epidermal Melanin Unit System] Das Epidermale Melanin-Einheit-System. *Dermatol. Wochenschr.,* **147**, 481-489.
34. Praetorius, C., Sturm, R.A. and Steingrimsson, E. (2014) Sun-induced freckling: ephelides and solar lentigines. *Pigment Cell Melanoma Res.*, **27**, 339-350.
35. Shlyueva, D., Stampfel, G. and Stark, A. (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.,* **15**, 272-286.
36. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA,* **107**, 21931-21936.
37. Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G. *et al.* (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.,* **41**, D56-63.
38. Strub, T., Giuliano, S., Ye, T., Bonet, C., Keime, C., Kobi, D., Le Gras, S., Cormont, M., Ballotti, R., Bertolotto, C. *et al.* (2011) Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene,* **30**, 2319-2332.
39. Levy, C., Khaled, M. and Fisher, D.E. (2006) MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol. Med.,* **12**, 406-414.
40. Li, J.Y., Song, J.S., Bell, R.J.A., Tran, T.N.T., Haq, R., Liu, H.F., Love, K.T., Langer, R., Anderson, D.G., Larue, L. *et al.* (2012) YY1 Regulates Melanocyte Development and Function by Cooperating with MITF. *PLoS Genet.,* **8**, e1002688.
41. Gorkin, D.U., Lee, D., Reed, X., Fletez-Brant, C., Bessling, S.L., Loftus, S.K., Beer, M.A., Pavan, W.J. and McCallion, A.S. (2012) Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Res.,* **22**, 2290-2301.

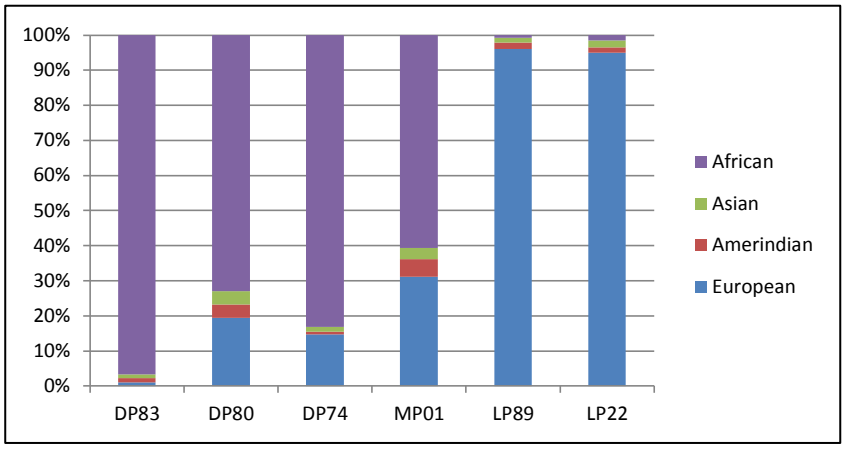
42. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.,* **15**, 1034-1050.
43. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature,* **491**, 56-65.
44. Eilbeck, K., Lewis, S.E., Mungall, C.J., Yandell, M., Stein, L., Durbin, R. and Ashburner, M. (2005) The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.,* **6**, R44.
45. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. *et al.* (2014) Ensembl 2014. *Nucleic Acids Res.,* **42**, D749-D755.
46. Felsenfeld, G. and Groudine, M. (2003) Controlling the double helix. *Nature,* **421**, 448-453.
47. McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y. and Pritchard, J.K. (2013) Identification of genetic variants that affect histone modifications in human cells. *Science*, **342**, 747-749.
48. Smale, S.T. and Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu. Rev. Biochem.,* **72**, 449-479.
49. Visel, A., Rubin, E.M. and Pennacchio, L.A. (2009) Genomic views of distant-acting enhancers. *Nature*, **461**, 199-205.
50. Zhang, X., Bailey, S.D. and Lupien, M. (2014) Laying a solid foundation for Manhattan - 'setting the functional basis for the post-GWAS era'. *Trends Genet.*, **30**, 140-149.
51. Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M. *et al.* (2014) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.,* **42**, D764-D770.
52. Marsman, J. and Horsfield, J.A. (2012) Long distance relationships: enhancer-promoter communication and dynamic gene transcription. *Biochim. Biophys. Acta*, **1819**, 1217-1227.
53. Bhoj, E.J., Ramos, P., Baker, L.A., Cost, N., Nordenskjold, A., Elder, F.F., Bleyl, S.B., Bowles, N.E., Arrington, C.B., Delhomme, B. *et al.* (2011) Human balanced translocation and mouse gene inactivation implicate Basonuclin 2 in distal urethral development. *Eur. J. Hum. Genet.,* **19**, 540-546.
54. Shpak, M., Hall, A.W., Goldberg, M.M., Derryberry, D.Z., Ni, Y., Iyer, V.R. and Cowperthwaite, M.C. (2014) An eQTL analysis of the human glioblastoma multiforme genome. *Genomics*, **103**, 252-263.
55. Song, H.L., Ramus, S.J., Tyrer, J., Bolton, K.L., Gentry-Maharaj, A., Wozniak, E., Anton-Culver, H., Chang-Claude, J., Cramer, D.W., DiCioccio, R. *et al.* (2009) A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat. Genet.,* **41**, 996-1000.
56. Wentzensen, N., Black, A., Jacobs, K., Yang, H.P., Berg, C.D., Caporaso, N., Peters, U., Ragard, L., Buys, S.S., Chanock, S. *et al.* (2011) Genetic Variation on 9p22 Is Associated with Abnormal Ovarian Ultrasound Results in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *PLoS One,* **6**, e21731.
57. Haraksingh, R.R. and Snyder, M.P. (2013) Impacts of variation in the human genome on gene regulation. *J. Mol. Biol.,* **425**, 3970-3977.
58. van den Boogaard, M., Barnett, P. and Christoffels, V.M. (2014) From GWAS to function: Genetic variation in sodium channel gene enhancer influences electrical patterning. *Trends Cardiovasc. Med.,* **24**, 99-104.
59. Visser, M., Kayser, M., Grosveld, F. and Palstra, R.J. (2014) Genetic variation in regulatory DNA elements: the case of OCA2 transcriptional regulation. *Pigment Cell Melanoma Res.,* **27**, 169-177.
60. Chardon, A., Cretois, I. and Hourseau, C. (1991) Skin colour typology and suntanning pathways. *Int. J. Cosmet. Sci.,* **13**, 191-208.
61. Del Bino, S., Sok, J., Bessac, E. and Bernerd, F. (2006) Relationship between skin response to ultraviolet exposure and skin color type. *Pigment Cell Res.,* **19**, 606-614.
62. Trost, A., Bauer, J.W., Lanschutzer, C., Laimer, M., Emberger, M., Hintner, H. and Onder, K. (2007) Rapid, high-quality and epidermal-specific isolation of RNA from human skin. *Exp. Dermatol.,* **16**, 185-190.
63. Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S.,Stadhouders, R., Palstra, R.J., Stevens, M., Kockx, C., van IJcken, W. (2010) The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.*, **24**, 277-289.

64. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.,* **14**, 178-192.
65. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics,* **24**, 2938-2939.
66. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. et al. (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.*, **28**, 1045-1048

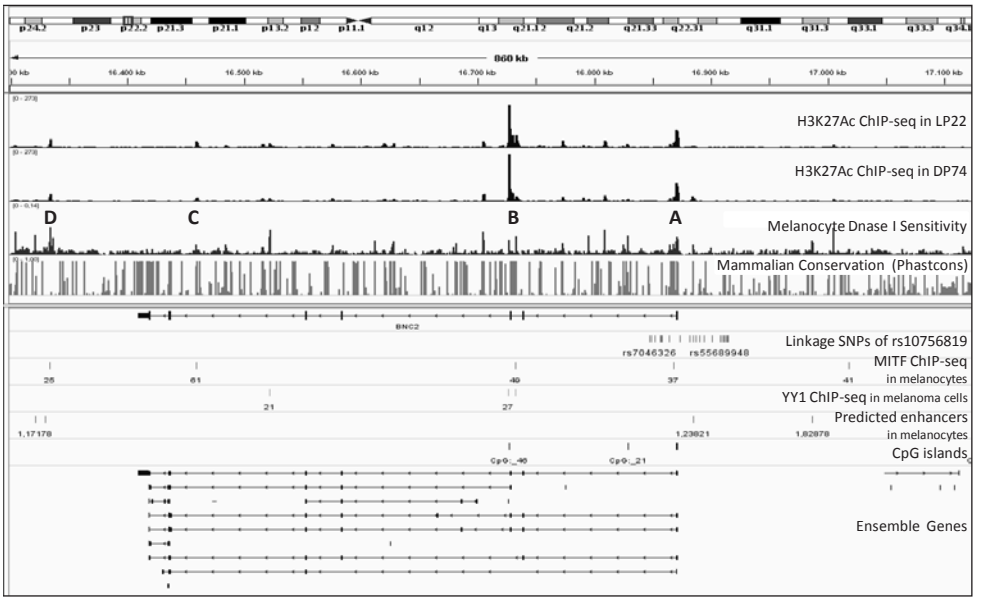
Supplemental figures



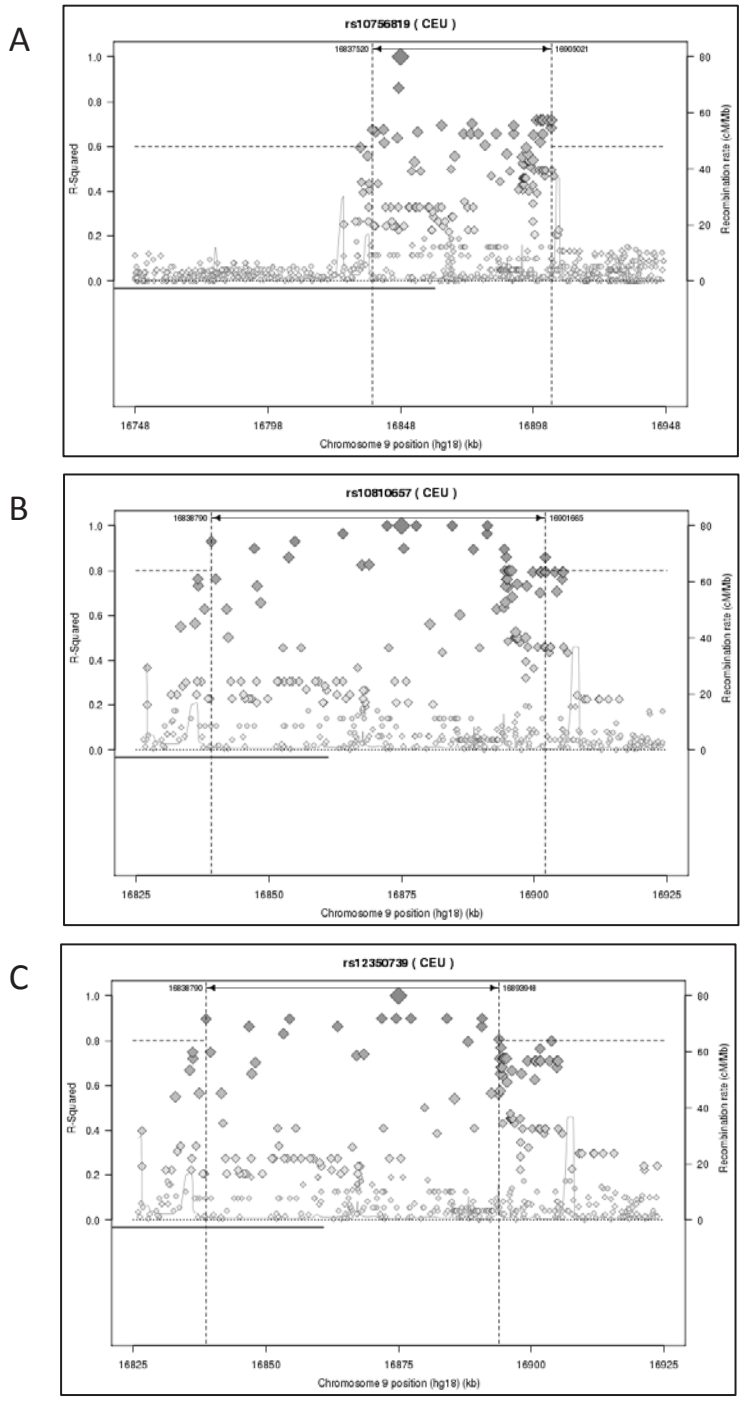
Supplemental Figure S1. Categorization of the skin epidermal samples. Graphical representation of the individual typology angle (ITA) values (49) of the epidermal skin samples shows that the 2 categories; light pigmented and dark pigmented, are completely separated.



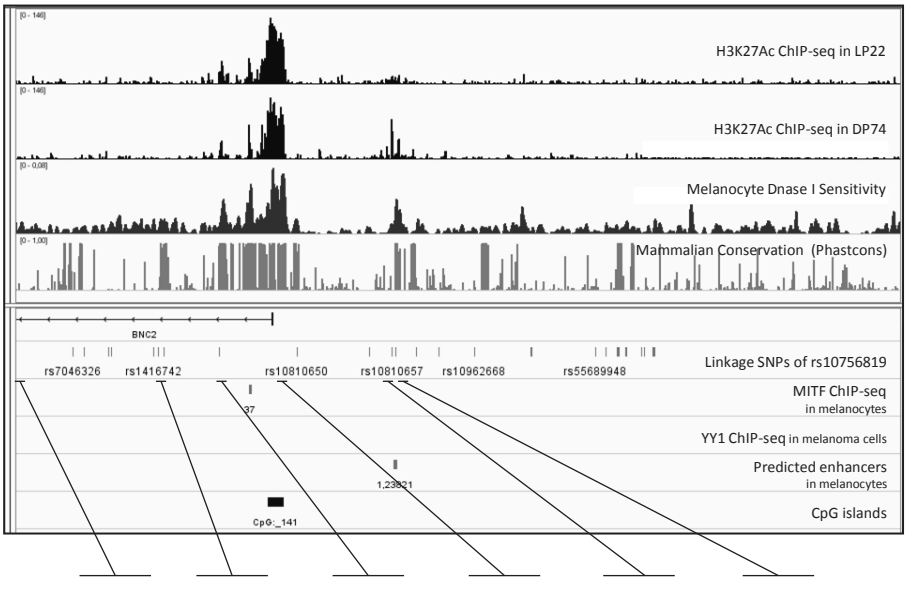
Supplemental Figure S2. Genetic ancestry analysis of the melanocyte cell lines and the skin epidermal samples. STRUCTURE (31) analysis of the 6 melanocyte cell lines using 24 autosomal ancestry-sensitive markers (ASMs) (32).

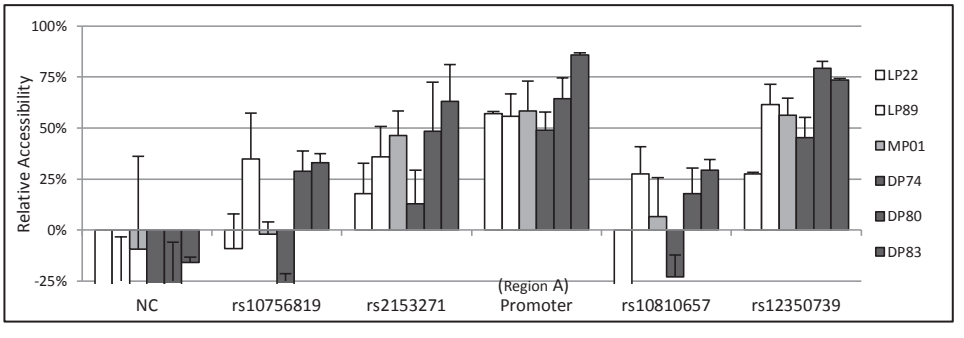


Supplemental Figure S3. Chromatin profile of the *BNC2* region. IGV genome browser (64) shows 860kb window of the *BNC2* region (hg coordinates chr9:16,298,464-17,160,559). To investigate the chromatin for features of enhancer elements, the following tracks are included: ChIP-seq analysis in LP22 and DP74 of acetylated histone H3 (H3K27Ac), an active chromatin mark (Palstra, manuscript in preparation); DNaseI hypersensitive sites in epidermal skin melanocytes ((37); Phastcons conserved elements inferred from 46 way alignments of placental mammals (42); SNPs in linkage disequilibrium (LD>0.6) with the pigmentation-associated SNP rs10756819 (see supplemental table S2a); ChIP-seq data for the transcription factor MITF in melanocytic cells (38); ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1 (40); predicted melanocyte-specific enhancers (41); CpG islands (37) and Ensemble transcripts (45). Based on the minimum criteria of the presence of a robust H3K27Ac peak and a DNaseI hypersensitivity signal, in combination with the presence of a MITF binding site, 4 regions with potential enhancer activity were selected, marked A-D. Note that the identified rs12350739 enhancer element is located 14kb upstream of the canonical *BNC2* promoter co-located with a predicted enhancer.

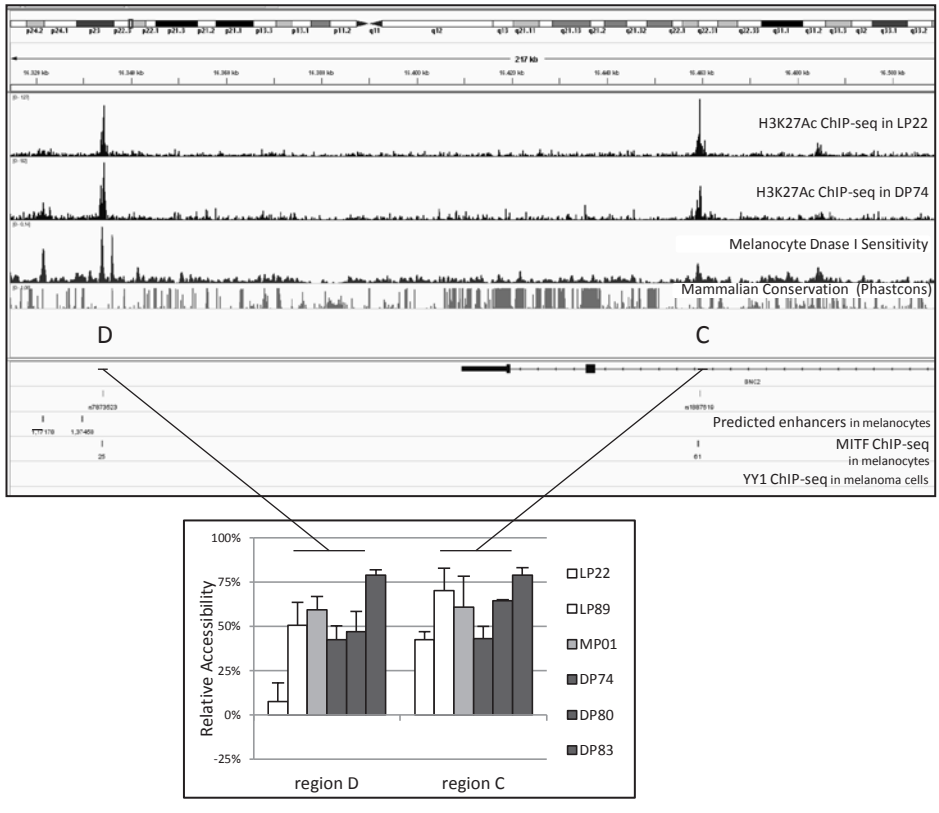


Supplemental Figure S4. Linkage disequilibrium plots of rs10756819, rs10810657 and rs12350739. **(A)** rs10756819; **(B)** rs10810657 and **(C)** rs12350739. Plots are generated with SNAP, a web-based tool (http://www.broadinstitute.org/mpg/snap/ (65)).

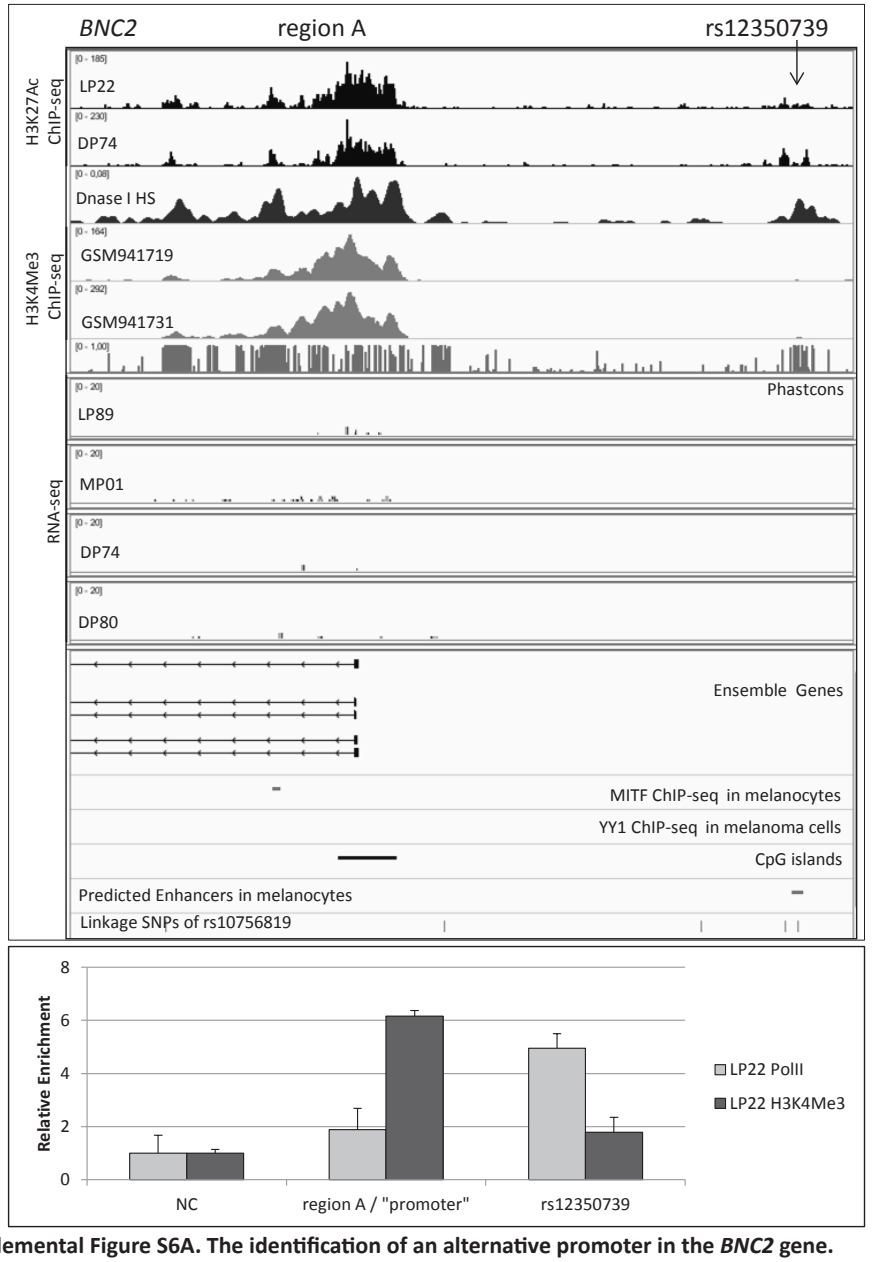




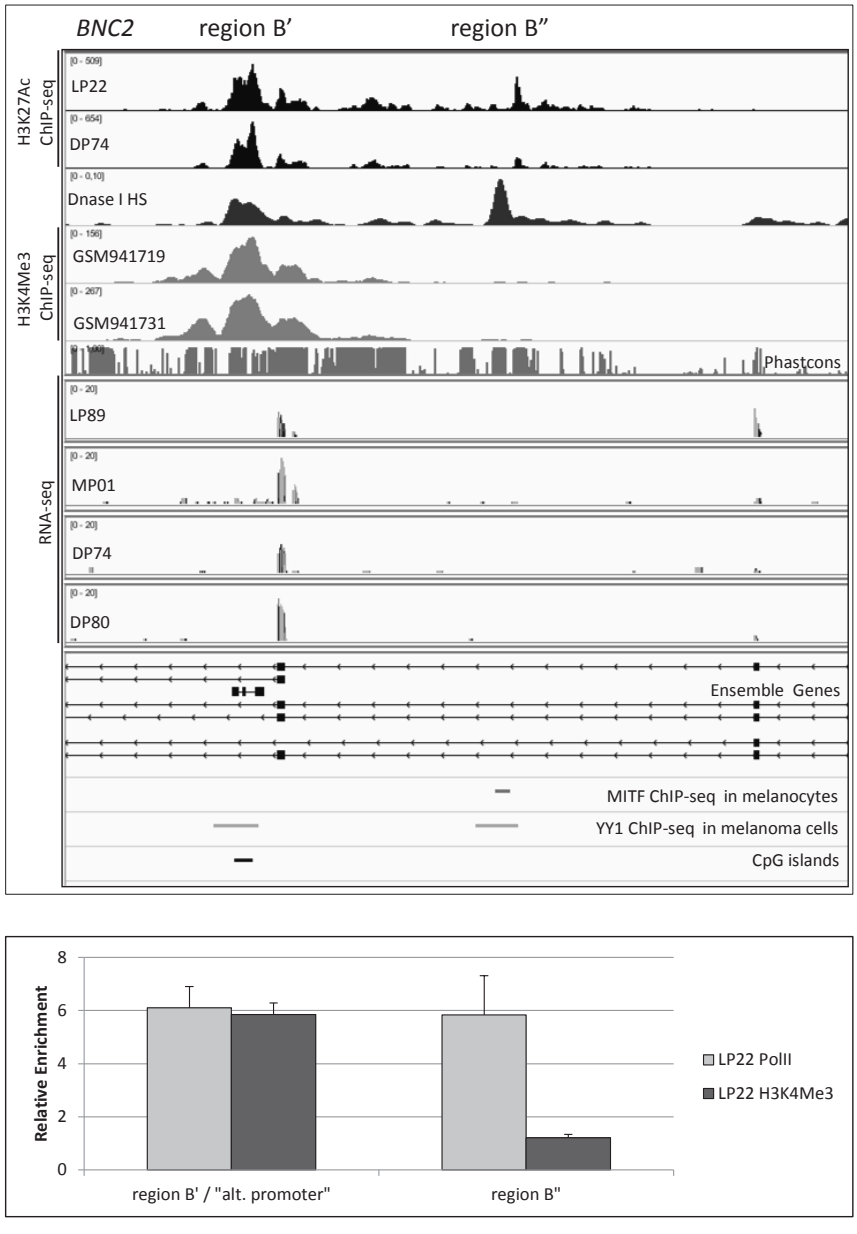
Supplemental Figure S5A. Additional regulatory elements within the *BNC2* locus.
IGV genome browser (64) shows (A) 100 kb window of the LD region containing the SNPs in linkage disequilibrium (LD > 0.6) with the pigmentation-associated SNP rs10756819 (hg coordinates chr9:16,840,877-16,943,644) and (B) 165 kb window of the *BNC2* region including regions marked C and D (see also supplemental figure S3). The following tracks are included: ChIP-seq analysis in LP22 and DP74 of acetylated histone H3 (H3K27Ac), an active chromatin mark (Palstra et al, manuscript in preparation); DNaseI hypersensitive sites in epidermal skin melanocytes (37); ChIP-seq data for the transcription factor MITF in melanocytic cells (38); ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1 (40); predicted melanocyte-specific enhancers (41) and Phastcons conserved elements inferred from 46 way alignments of placental mammals (42).
(A) Results of the EpiQ analyses at six different locations in the LD region; the chromatin is inaccessible at the (negative control) region NC, but also at the regions around rs10756819 and at rs10810657, it is moderately accessible at the region around rs2153271. The chromatin is open and accessible at the canonical promoter of *BNC2* in region A and at the identified enhancer element surrounding rs12350739.



Supplemental Figure S5B. Additional regulatory elements within the *BNC2* locus. (B) Results of the EpiQ analyses at regions C and D shows moderately accessible chromatin in both regions.



Supplemental Figure S6A. The identification of an alternative promoter in the *BNC2* gene.
IGV genome browser (64) shows the regions around (A) the enhancer region at rs12350739 and the canonical promoter at exon 1 in *BNC2* and (B) exon 2 and 3 of *BNC2*. To test for the presence of an alternative promoter at the exon 3 region, two H3K4Me3 ChIP-seq datasets in human skin melanocytes (66) and RNA-seq data of 4 cell lines (LP9, DP80, DP74 and MP01) were investigated. Number of transcripts are indicated by peak heights and grey bars. Note that at exon 1 low levels of transcripts are present, whereas at exon 3 transcripts are present at higher levels. Results of the H3K4Me3 and RNA polymerase II ChIP-qPCR analyses at five different locations in the *BNC2* region; H3K4Me3 ChIP-qPCR confirms the presence of the canonical promoter at exon 1 **(A),** *(continued on next page)*



Supplemental Figure S6B. The identification of an alternative promoter in the *BNC2* gene.
and the alternative promoter at exon 3 **(B)**, while RNA polymerase II binding confirms the differences of transcription levels between exon1-2 and exon3-4. In addition, RNA polymerase II binding is also detected at the rs12350739 enhancer element and at region B".

Supplemental Tables

Supplemental Table S1

	HIrisPlex results					
HAIR	**LP22**	**LP89**	**MP01**	**DP74**	**DP80**	**DP83**
Brown	0,119	0,314	0,227	0,168	0,288	0,101
Red	0,005	0,005	0,000	0,000	0,000	0,000
Black	0,050	0,422	0,371	0,828	0,651	0,896
Blond	0,826	0,259	0,402	0,003	0,061	0,003
SHADE						
Light	0,948	0,227	0,724	0,004	0,063	0,002
Dark	0,052	0,773	0,276	0,996	0,937	0,998
HAIR color result	**blond**	**(dark) brown**	**(light) brown**	**black**	**dark brown/ black**	**black**
EYE						
Blue	0,899	0,207	0,457	0,000	0,024	0,000
Int.	0,066	0,161	0,162	0,025	0,083	0,006
Brown	0,035	0,632	0,381	0,975	0,892	0,994
EYE color result	**blue**	**brown**	**blue/int**	**brown**	**brown**	**brown**

HIrisPlex results of the 6 melanocyte cell lines.
HIrisPlex analysis (29) predicts the light pigmented cell line LP22 to have blue eyes and light-shaded blond hair, whereas the other light pigmented cell donor most probably has dark-brown hair with brown eyes. The MP01 cell donor is predicted to have blond hair with blue eyes, and the 3 dark pigmented cell donors of DP74, DP80 and DP83 all have dark-brown to black hair, and brown eyes. These predicts confirm the phenotypic information provided by the cell line supplier (Cascade Biologics).

Supplemental Table S2

sample number	Questionnaire Data			skin color ITA	HIrisPlex HAIR color and Shade						HIrisPlex HAIR color result	HIrisPlex EYE color			
	geographical ancestry	hair color	eye color	category	Brown	Red	Black	Blond	Light Shade	Dark Shade		Blue	Int.	Brown	Eye color result
1	Surinam	dark brown/black	brown	dark	0,093	0,000	0,903	0,004	0,002	0,998	black	0,000	0,006	0,994	brown
2	Netherlands	grey	green	light	0,168	0,004	0,149	0,679	0,708	0,292	dark blond	0,899	0,066	0,035	blue
3	Surinam	dark brown/black	dark brown/black	dark	0,997	0,000	0,000	0,003	1,000	0,000	light brown	0,000	0,006	0,994	brown
4	Netherlands	dark blond (grey)	blue	light	0,308	0,003	0,207	0,482	0,604	0,396	dark blond/brown	0,943	0,045	0,012	blue
5	Netherlands	dark blond	blue	light	0,240	0,159	0,149	0,451	0,699	0,301	dark blond/brown	0,943	0,045	0,012	blue
6	Netherlands	dark blond (grey)	blue-green-grey	light	0,143	0,087	0,105	0,666	0,815	0,185	dark blond	0,870	0,076	0,053	blue
7	Curacao	dark brown/black	brown	dark	0,124	0,000	0,873	0,003	0,002	0,998	black	0,000	0,006	0,994	brown
8	Netherlands	brown (grey)	brown	light	0,278	0,002	0,712	0,009	0,010	0,990	black	0,280	0,315	0,405	int/brown
9	Netherlands	dark brown/black	brown	light	0,323	0,014	0,516	0,147	0,170	0,830	brown	0,024	0,083	0,892	brown
10	Netherlands	dark blond (grey)	green	light	0,187	0,007	0,166	0,641	0,730	0,270	dark blond	0,870	0,076	0,053	blue
11	Netherlands	dark blond	blue	light	0,122	0,069	0,030	0,779	0,924	0,076	blond	0,962	0,025	0,013	blue
12	Surinam	dark brown/black	dark brown/black	dark	0,116	0,000	0,883	0,002	0,001	0,999	black	0,000	0,006	0,994	brown
13	Netherlands	dark blond	blue (grey)	light	0,071	0,072	0,019	0,838	0,958	0,042	blond	0,870	0,076	0,053	blue
14	Surinam	dark brown	dark brown	dark	0,408	0,002	0,462	0,129	0,089	0,911	dark brown	0,001	0,026	0,973	brown
15	Netherlands	brown	brown	light	0,282	0,004	0,470	0,244	0,240	0,760	dark brown	0,207	0,161	0,632	brown
16	Netherlands	blond	blue	light	0,423	0,010	0,252	0,315	0,478	0,522	light brown	0,945	0,049	0,006	blue
17	Netherlands	dark blond	grey/green	light	0,232	0,117	0,059	0,592	0,879	0,121	dark blond	0,899	0,066	0,035	blue
18	Netherlands	blond	green	light	0,158	0,011	0,067	0,764	0,908	0,092	blond	0,919	0,048	0,033	blue
19	Netherlands	dark blond (grey)	grey/green	light	0,234	0,040	0,176	0,550	0,672	0,328	dark blond	0,503	0,211	0,286	blue/int
20	Eritrea	dark brown	brown	dark	0,124	0,000	0,871	0,004	0,005	0,995	black	0,000	0,006	0,994	brown
21	Netherlands	blond	blue (green)	light	0,090	0,013	0,164	0,733	0,816	0,184	blond	0,951	0,034	0,014	blue
22	Surinam	brown (copper)	brown	dark	0,184	0,000	0,795	0,020	0,020	0,980	black	0,000	0,013	0,987	brown
23	Surinam	dark brown/black	dark brown	dark	0,147	0,000	0,843	0,010	0,012	0,988	black	0,000	0,006	0,994	brown
24	Surinam	black (dark brown)	brown	dark	0,171	0,000	0,806	0,023	0,015	0,985	black	0,009	0,043	0,948	brown
25	Curacao	dark brown	dark brown	dark	0,126	0,000	0,866	0,008	0,011	0,989	black	0,000	0,008	0,992	brown
26	Surinam	black (dark brown)	brown	dark	0,114	0,000	0,882	0,004	0,003	0,997	black	0,000	0,006	0,994	brown
27	Curacao	brown	brown	dark	0,083	0,000	0,913	0,003	0,002	0,998	black	0,000	0,006	0,994	brown
28	Netherlands	dark blond	green	light	0,184	0,012	0,148	0,655	0,723	0,277	dark blond	0,919	0,048	0,033	blue
29	Netherlands	brown	brown	light	0,212	0,052	0,228	0,508	0,575	0,425	dark blond	0,229	0,128	0,643	brown

Phenotype and geographical ancestry information of skin epidermal samples.

Geographical ancestry information is based on place of birth of the sample donor and that of their parents and grandparents. The self-reported hair and eye color information is confirmed with HIrisPlex analysis. Skin color is categorized according to the ITA values as is described in the materials and methods section, and displayed in supplemental figure S1.

Supplemental table S3.

A

linked SNPs of rs10756819								
SNP	Proxy	Distance	RSquared	DPrime	Chromosome	start	end	
rs10756819	rs10756819	0	1.000	1.000	chr9	16858083	16858084	
rs10756819	rs10962643	681	0.861	1.000	chr9	16857402	16857403	
rs10756819	rs10962684	51249	0.717	0.879	chr9	16909332	16909333	
rs10756819	rs10738468	52679	0.717	0.879	chr9	16910762	16910763	
rs10756819	rs7045767	52814	0.717	0.879	chr9	16910897	16910898	
rs10756819	rs4366169	53554	0.717	0.879	chr9	16911637	16911638	
rs10756819	rs4445329	53673	0.717	0.879	chr9	16911756	16911757	
rs10756819	rs6475092	55389	0.717	0.879	chr9	16913472	16913473	
rs10756819	rs7032221	56811	0.717	0.879	chr9	16914894	16914895	
rs10756819	rs3814113	56937	0.717	0.879	chr9	16915020	16915021	
rs10756819	rs74664507	55752	0.714	0.956	chr9	16913835	16913836	
rs10756819	**rs12350739**	**26933**	**0.702**	**0.916**	**chr9**	**16885016**	**16885017**	
rs10756819	rs10810650	15467	0.694	0.879	chr9	16873550	16873551	
rs10756819	rs28498684	42611	0.693	0.879	chr9	16900694	16900695	
rs10756819	rs10810671	56751	0.683	0.874	chr9	16914834	16914835	
rs10756819	rs4961501	6406	0.675	1.000	chr9	16851677	16851678	
rs10756819	rs7046326	10564	0.675	1.000	chr9	16847519	16847520	
rs10756819	rs2153271	6437	0.665	0.876	chr9	16864520	16864521	
rs10756819	rs1339552	9294	0.665	0.876	chr9	16848789	16848790	
rs10756819	rs12345776	23793	0.657	0.841	chr9	16881876	16881877	
rs10756819	**rs10810657**	**26502**	**0.657**	**0.841**	**chr9**	**16884585**	**16884586**	
rs10756819	rs3927680	29282	0.657	0.841	chr9	16887365	16887366	
rs10756819	rs10962668	36056	0.657	0.841	chr9	16894139	16894140	
rs10756819	rs36116821	42681	0.657	0.841	chr9	16900764	16900765	
rs10756819	rs10465044	53581	0.656	0.840	chr9	16911664	16911665	
rs10756819	rs55689948	50085	0.651	0.870	chr9	16908168	16908169	
rs10756819	rs1416742	1201	0.638	0.873	chr9	16856882	16856883	
rs10756819	rs10738467	52593	0.619	0.865	chr9	16910676	16910677	
rs10756819	rs7868157	6107	0.616	0.830	chr9	16851976	16851977	
rs10756819	rs10962662	31853	0.605	0.946	chr9	16889936	16889937	

B

linked SNPs rs10810657							
SNP	Proxy	Distance	RSquared	DPrime	start	end	
rs10810657	rs10810657	0	1.000	1.000	16884585	16884586	
rs10810657	rs12345776	2709	1.000	1.000	16881876	16881877	
rs10810657	rs3927680	2780	1.000	1.000	16887365	16887366	
rs10810657	rs10962668	9554	1.000	1.000	16894139	16894140	
rs10810657	rs36116821	16179	1.000	1.000	16900764	16900765	
rs10810657	rs10810650	11035	0.965	1.000	16873550	16873551	
rs10810657	rs28498684	16109	0.965	1.000	16900694	16900695	
rs10810657	rs2153271	20065	0.931	1.000	16864520	16864521	
rs10810657	rs1339552	35796	0.931	1.000	16848789	16848790	
rs10810657	**rs12350739**	**431**	**0.899**	**1.000**	**16885016**	**16885017**	
rs10810657	rs1416742	27703	0.899	1.000	16856882	16856883	
rs10810657	rs7029285	19362	0.896	1.000	16903947	16903948	
rs10810657	rs10962672	13533	0.895	0.963	16898118	16898119	
rs10810657	rs58691828	19717	0.860	0.961	16904302	16904303	
rs10810657	rs62541919	21222	0.859	0.927	16863363	16863364	
rs10810657	rs10465044	27079	0.859	0.927	16911664	16911665	
rs10810657	rs10122763	6093	0.827	0.960	16878492	16878493	
rs10810657	rs10810655	7448	0.825	0.925	16877137	16877138	
rs10810657	rs7032175	19494	0.800	1.000	16904079	16904080	
rs10810657	rs7033194	20119	0.800	1.000	16904704	16904705	
rs10810657	rs7033354	20260	0.800	1.000	16904845	16904846	
rs10810657	rs10756835	20362	0.800	1.000	16904947	16904948	
rs10810657	rs12344726	20742	0.800	1.000	16905327	16905328	

C

linked SNPs rs12350739							
SNP	Proxy	Distance	RSquared	DPrime	start	end	
rs12350739	rs12350739	0	1.000	1.000	16875017	16875018	
rs12350739	**rs10810657**	**431**	**0.899**	1.000	16874586	16874587	
rs12350739	rs3927680	2349	0.899	1.000	16877366	16877367	
rs12350739	rs12345776	3140	0.899	1.000	16871877	16871878	
rs12350739	rs10962668	9123	0.899	1.000	16884140	16884141	
rs12350739	rs36116821	15748	0.899	1.000	16890765	16890766	
rs12350739	rs2153271	20496	0.897	0.964	16854521	16854522	
rs12350739	rs1339552	36227	0.897	0.964	16838790	16838791	
rs12350739	rs10810650	11466	0.863	0.963	16863551	16863552	
rs12350739	rs28498684	15678	0.863	0.962	16890695	16890696	
rs12350739	rs1416742	28134	0.863	0.929	16846883	16846884	
rs12350739	rs62541919	21653	0.831	0.961	16853364	16853365	
rs12350739	rs7029285	18931	0.806	1.000	16893948	16893949	

SNPs in linkage disequilibrium with rs10756819, rs10810657 and rs12350739. Data is generated with SNAP, a web-based tool (http://www.broadinstitute.org/mpg/snap/ (65)).
(A) rs10756819 (r2 > 0.6)
(B) rs10810657 (r2 > 0.8)
(C) rs12350739 (r2 > 0.8)

Supplemental table S4.

		BNC2 genotypes		
	phenotype	rs10756819	rs10810657	rs12350739
1	dark	A	T	G
2	light	A	A	A
3	dark	G	T	G
4	light	A	A	A
5	light	AG	AT	AG
6	light	AG	AT	AG
7	dark	G	AT	G
8	light	A	A	A
9	light	A	A	A
10	light	AG	T	G
11	light	A	AT	AG
12	dark	G	T	G
13	light	AG	T	G
14	dark	AG	T	G
15	light	A	AT	AG
16	light	A	A	A
17	light	A	A	A
18	light	AG	A	A
19	light	AG	T	AG
20	dark	AG	AT	AG
21	light	AG	AT	AG
22	dark	AG	T	G
23	dark	G	AT	G
24	dark	AG	AT	A
25	dark	G	AT	G
26	dark	AG	T	G
27	dark	G	T	G
28	light	AG	AT	AG
29	light	A	A	A

Genotypes of rs10756819, rs10810657 and rs12350739 in the skin epidermal samples.

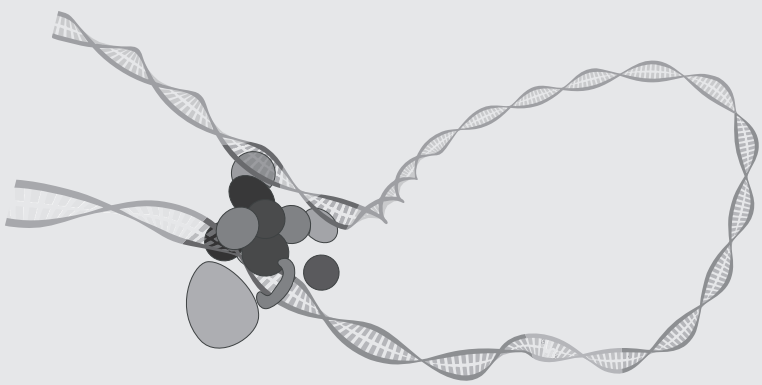
Supplemental Table S5 is available at the Human Molecular Genetics website.

Chapter 4

Allele-specific transcriptional regulation of *IRF4* in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the *IRF4* promoter

Mijke Visser, Robert-Jan Palstra, and Manfred Kayser

Human Molecular Genetics, in press (2015)



Abstract

The majority of significant single-nucleotide polymorphisms (SNPs) identified with genome-wide association studies are located in non-coding regions of the genome; it is therefore possible that they are involved in transcriptional regulation of a nearby gene rather than affecting an encoded protein's function. Previously it was demonstrated that the SNP rs12203592, located in intron 4 of the *IRF4* gene, is strongly associated with human skin pigmentation and modulates an enhancer element that controls expression of *IRF4*. In our study, we investigated the allele-specific effect of rs12203592 on *IRF4* expression in epidermal skin samples and in melanocytic cells from donors of different skin color. We focused on the characteristics and activity of the enhancer, and on long-range chromatin interactions in melanocytic cells homozygous and heterozygous for rs12203592. We found that, irrespective of the trans-activating environment, *IRF4* transcription is strongly correlated with the allelic status of rs12203592, the activity of the rs12203592 enhancer and that the chromatin features depend on the rs12203592 genotype. Furthermore, we demonstrate that the rs12203592 enhancer physically interacts with the *IRF4* promoter through an allele-dependent chromatin loop, and suggest that subsequent allele-specific activation of *IRF4* transcription is stabilized by another allele-specific loop from the rs12203592 enhancer to an additional regulatory element in *IRF4*. We conclude that the non-coding SNP rs12203592 is located in a regulatory region and affects a wide range of enhancer characteristics, resulting into modulation of the enhancer's activity, its interaction with the *IRF4* promoter, and subsequent allele-specific transcription of *IRF4*. Our findings provide another example of a non-coding SNP affecting skin color by modulating enhancer-mediated transcriptional regulation.

Introduction

Single nucleotide polymorphisms (SNPs) linked to specific phenotypes in genome-wide association studies (GWA studies) are often located in non-coding regions of the genome. In contrast to coding variants affecting protein function, these non-coding SNPs are more complicated to functionally characterize (1). The SNP rs12203592 is located in intron 4 of the interferon regulatory factor 4 (*IRF4*) gene, encoding a member of a helix-loop-helix family of DNA-binding transcription factors involved in downstream regulation of interferon signaling. IRFs are primarily associated with immune system development and response (2). *IRF4* is predominantly expressed in lymphocytes, macrophages, B cells and dendritic cells, but also in melanocytic lineages (3). Involvement of this gene has been described in several types of lymphoma and leukemia (3), including childhood acute lymphoblastic leukemia (ALL) (4), but also in non-hematopoietic diseases (3). *IRF4* was initially implicated in pigmentation through the association of DNA variants located outside this gene, particularly with freckling in humans (5). In succession the intronic SNP rs12203592 was found to be strongly associated with hair, eye, skin color, tanning response and nevus count (6–10). This non-coding SNP was also identified to be associated with (male-specific) childhood ALL and in that study the first evidence was given that rs12203592 is located within a regulatory element, controlling expression of *IRF4* through binding of transcription factor TFAP2α (4). In Burkitt Lymphoma

B-cells (Raji), HEK293T human embryonic kidney cells and in NCI-H295R human adrenal cells, the rs12203592 region containing the C-allele was shown to strongly bind the transcription factor TFAP2α and repress *IRF4* promoter activity, while the rs12203592 region with the T-allele binds TFAP2α with less affinity, and has a less repressive effect on *IRF4* expression (4). More recently it was shown that in skin melanocytes, rs12203592 is located within a melanocyte-specific enhancer regulating the expression of *IRF4* (11). The transcription factor TFAP2α binds to the enhancer element containing the C-allele with high affinity, and together with MITF, transcription of *IRF4* is activated. When the rs12203592 T-allele is present, TFAP2α-binding is reduced, which in turn leads to a decreased expression of *IRF4* (11). In addition, the lighter pigmentation phenotype associated with the rs12203592 T-allele was shown to result from impaired activation of tyrosinase (*TYR*) transcription due to reduced *IRF4* levels (11). The difference between the two studies, describing either transcriptional repression (4) or activation (11) might be due to cell-type-specific effects mediated by transcription factors other than TFAP2α and to the fact that TFAP2α can behave as an activator as well as a repressor in different contexts (12, 13).

In the present study we extend on these previously published results (4, 11) and provide detailed novel insight in the molecular mechanism of *IRF4* activation by the rs12203592 enhancer. Using several molecular approaches, we investigated the transcriptional regulation of *IRF4* in melanocytes, with a special focus on allelic differences of the rs12203592 enhancer. We studied the expression of *IRF4* and transcription factors potentially involved in its transcriptional regulation in epidermal skin samples from donors of different skin color phenotypes and rs12203592 genotypes, in cultured skin melanocytes derived from differently pigmented skin as well as in two melanoma cell lines. We surveyed the chromatin profile of the *IRF4* locus focusing at specific regulatory sites within the gene using chromatin immune-precipitation (ChIP) of H3K27Ac in combination with deep sequencing (ChIP-seq) and the use of ENCODE data, and we mapped the chromatin folding of the *IRF4* locus in cultured skin melanocytes and melanoma cell lines using chromosome conformation capture (3C) techniques. We demonstrate an allele-dependent activity of the rs12203592 enhancer, and subsequent differential *IRF4* expression. Furthermore, we show that the rs12203592 enhancer physically interacts with the *IRF4* promoter and we provide evidence for an interaction with a potential additional regulatory element in intron 7 of *IRF4*, both interactions depend on the rs12203592 genotype. We propose a model for the transcriptional regulation of *IRF4*, in which we suggest that the activation of *IRF4* transcription in melanocytes is determined by the rs12203592 enhancer/*IRF4* promoter chromatin-loop, while the potential rs12203592 enhancer/intron-7 regulatory-element chromatin-loop mediates stabilization of the chromatin structure resulting into an increased and stable transcription of *IRF4*.

Results

Expression levels of IRF4 depend on the allelic status of rs12203592 in epidermal skin samples.

To study the correlation between the rs12203592 genotype and *IRF4* gene expression levels, we used a panel of 29 epidermal skin samples obtained from donors with different pigmentation phenotypes (12 dark and 17 light) and three commercially available primary Human Epidermal Melanocytes of neonatal origin derived from donors with different pigmentation phenotypes; one light pigmented donor (LP22) and two dark pigmented donors (DP74 and DP80) (see 14 for category separation of the epidermal skin samples and characterization of all samples).

DNA and RNA were co-extracted from the epidermal layers of the skin biopsy samples and mRNA levels of *IRF4* were measured in the individual samples using quantitative (q) reverse transcriptase (RT)-PCR. DNA samples were used for HIrisPlex analysis, and as rs12203592 is included in this assay, the genotypes of rs12203592 were obtained from

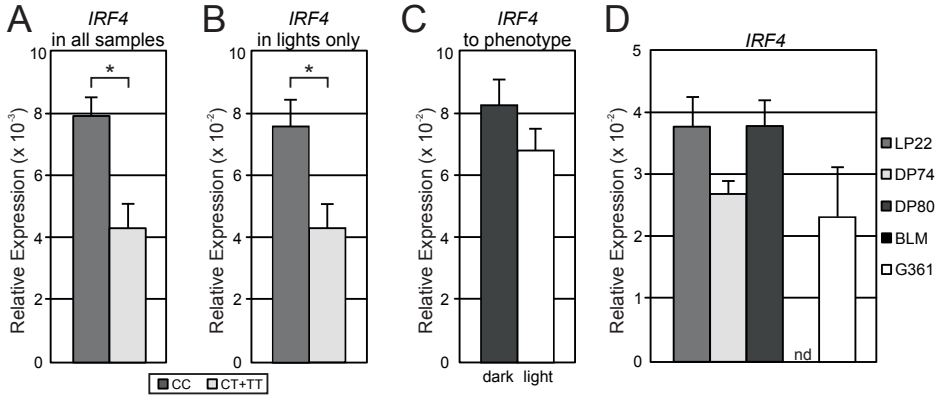


Figure 1. *IRF4* expression levels depend on the allelic status of rs12203592. RT-qPCR analysis of *IRF4* transcripts (A) in all epidermal skin samples and (B) in only light epidermal skin samples with either the rs12203592 CC-genotype (all: n=25; lights: n=13) or with the combinational rs12203592 CT- and TT-genotypes (N=4) demonstrates reduced *IRF4* expression when the T-allele is present. (C) RT-qPCR analysis of *IRF4* transcripts in all epidermal skin samples with either dark or light skin color phenotype indicates no differential expression of *IRF4* between the two pigmentation phenotype categories. (D) RT-qPCR analysis of *IRF4* transcripts in three melanocyte cell lines and two melanoma cell lines. *IRF4* is differentially expressed between the two cell lines with the rs12203592 CC-genotype (LP22 and DP80) and the cell line with the rs12203592 CT-genotype (DP74), *IRF4* expression is not detected in the BLM melanoma cell line, while expression of *IRF4* is detected at slightly decreased levels in the G361 melanoma cell lines with the rs12203592 CC-genotype when compared with the melanocyte cell lines with similar rs12203592 genotypes. Each individual gene-expression analysis is performed in triplicate and normalized to an endogenous control (*ACTB*). Averaged expression values for the genotype (CC; CT+TT) and skin color phenotype (light; dark) categories of the epidermal skin samples are calculated after normalization with *ACTB*. Data are represented as mean ± SEM; * $P < 0.05$.

the HIrisPlex results for all epidermal skin samples (14, 15). Analyzing this SNP in a worldwide set of individuals from 51 populations (HGDP-CEPH) (16) reveals that the minor-allele-frequency of this SNP is relatively low; it is present in 12% of the average European populations and it is rare in African population (0.6%) and in Asian populations (1%). Similar results are obtained when analyzing data from the 1000 Genomes project (17). This is reflected in the rs12203592 genotypes of the epidermal skin samples we used in this study; only 1 out of 29 epidermal skin sample is homozygote TT, and 3 epidermal skin samples are heterozygote CT, all 4 are of European decent (Supplementary Material, Table S1).

We observed statistically significant higher expression levels of *IRF4* in the samples carrying the homozygote CC-genotype (n=23) than in the combined sample set with the heterozygote CT- and homozygote TT-genotypes (n=4) (Fig. 1A, p=0.02). This confirms previous notions that transcription of *IRF4* depends on the allelic status of rs12203592 (11). To exclude the possibility that the detected differential *IRF4* expression between CC and CT/TT samples is caused by the dark phenotype of a large fraction (48%) of the rs12203592-CC epidermal skin samples, we analyzed the correlation between *IRF4* expression and the rs12203592 genotype in the light sample set only. This analysis revealed that *IRF4* is still differentially expressed between rs12203592 CC and CT/TT alleles when only the light samples are considered, although the difference is less significant (p=0.048) (Fig. 1B), which is most probably due to the smaller sample size (n=17). Furthermore, we observed no significant expression differences for *IRF4* between the light and dark epidermal skin samples (Fig. 1C).

We next investigated the expression patterns of *IRF4* in the melanocyte cell lines LP22, DP74 and DP80. DNA and RNA were co-extracted from the individual cell lines, mRNA levels were measured using RT-qPCR and DNA samples were used to obtain rs12203592 genotypes using the HIrisPlex (14, 15). Interestingly, this revealed that not a lightly pigmented but the darkly pigmented melanocytic cell line DP74 is heterozygous for rs12203592 (Supplementary Material, Table S1). Despite the dark phenotype of DP74, expression of *IRF4* is lower in this cell line when compared with the expression of *IRF4* in LP22 and DP80 (Fig. 1D), which are both homozygote CC for rs12203592 (Supplementary Material, Table S1).

From the results with both samples sets we conclude that *IRF4* is differentially expressed depending on the allelic status of rs12203592, confirming previous observations (11) and extending on this knowledge, we demonstrate that this differential expression of *IRF4* does not depend on the pigmentation phenotype of the epidermal skin samples and the melanocytic cell lines used.

Characterization of intronic regulatory elements in IRF4.

The region around rs12203592 was shown to act as an intronic enhancer for *IRF4* in melanocytes (11), and we demonstrated in the previous paragraph that the variation in *IRF4* transcription levels strongly depends on the allelic status of this SNP. Regulatory elements in the genome are characterized by several features (18), including histone modifications, chromatin accessibility, transcription factor binding and evolutionary conservation. To study in more detail the mode-of-action of the rs12203592 enhancer, we profiled the chromatin of the *IRF4* region for the presence of acetylated histone H3 (H3K27Ac), which is an active

chromatin mark (19), using ChIP-seq analyses in the homozygote-CC melanocytic cell line LP22 and the heterozygote-CT melanocytic cell line DP74 (Palstra et al manuscript in preparation). These data were combined with previously published data sets for DNaseI hypersensitive sites (DHSs) in epidermal skin melanocytes (20), ChIP-seq data for the melanocyte master regulator MITF in melanocytic cells (21), ChIP-seq data in MALME-3M melanoma cells for the more ubiquitously expressed and MITF interacting transcription factor YY1 (22), and Phastcons conserved elements inferred from 46 way alignments of placental mammals (23). The locus-wide profile of *IRF4* revealed two regions that are clearly enriched for the active enhancer mark H3K27Ac; the *IRF4* promoter region and the region around rs12203592 (Fig. 2A). When zooming into this highly conserved region around

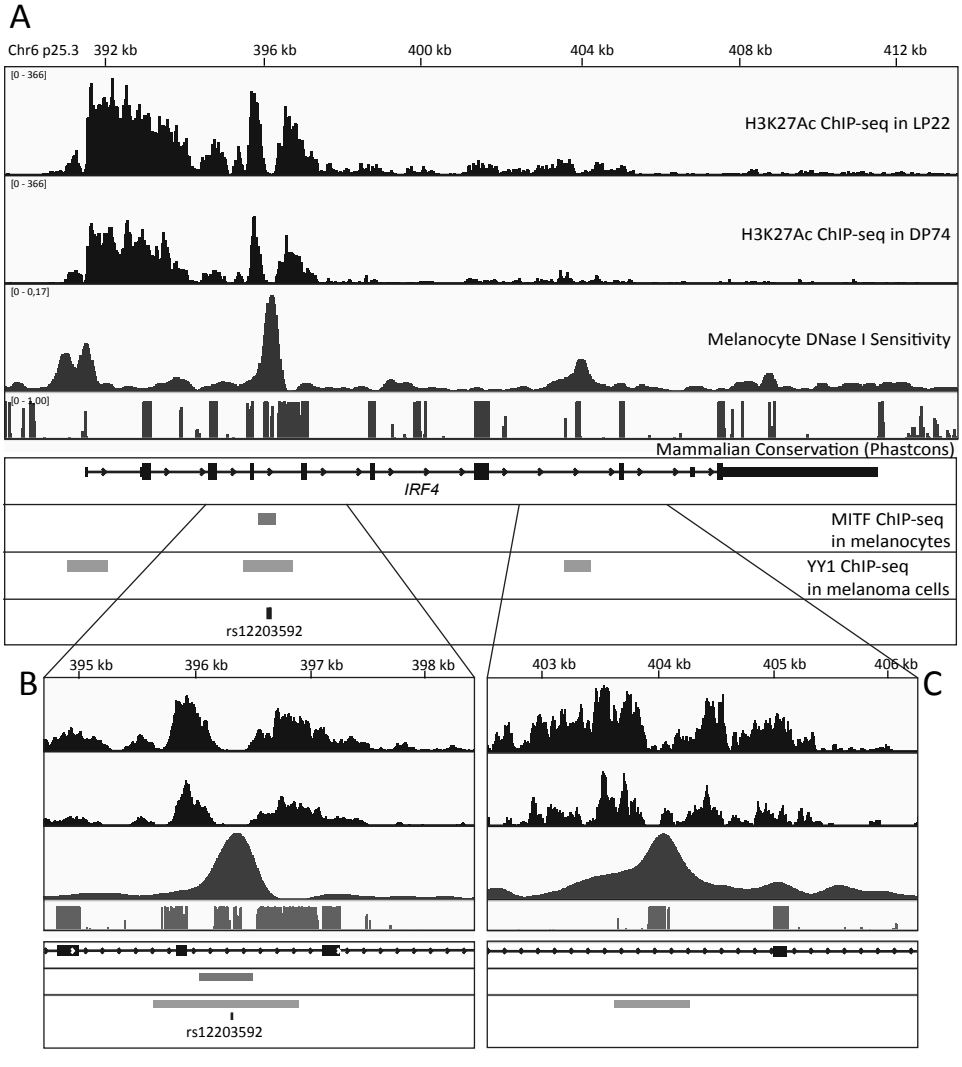


Figure 2. *(legend at the bottom of the next page)*

rs12203592, we observed that the H3K27Ac enrichment signals are differential between LP22 and DP74; the signals are more prominent in the LP22 cells carrying the CC-genotype than in the DP74 cells with the CT-genotype (Fig. 2B). To exclude the possibility that these differential H3K27Ac signals are due to experimental factors, we investigated H3K27Ac enrichment at the neighboring *DUSP22* gene, which has no known function in human pigmentation biology. We found that in this region the H3K27Ac signals are not differential between LP22 and DP74 (Supplementary Material, Fig. S1). Moreover, we reported higher H3K27Ac signals in DP74 than in LP22 at the intergenic enhancer and at the alternative promoter of the pigmentation-associated *BNC2* gene while the canonical promoter of *BNC2* displays equal H3K27Ac enrichment in these cell lines (14). Taken together, we conclude from these observations that the differences in H3K27Ac enrichment observed in the *IRF4* region are not due to technical aspects, but rather to the different rs12203592 genotypes.

Accessibility of chromatin is another important characteristic of enhancer elements – regions that are active, are generally more open and accessible for protein complexes like the transcription machinery to bind – and this can be detected by several techniques. DHS mapping detects regions of chromatin that are more accessible for the DNA digesting activity of the DNase I enzyme. In a publicly available dataset for DHS signals obtained from epidermal skin melanocytes (20) such a prominent DHS signal is present at the rs12203592 enhancer (Fig. 2B). To investigate this in our melanocytic cell system, we used a commercially available assay that directly measures the accessibility of chromatin to nucleases (EpiQ-assay; BioRad). This revealed that the chromatin at the rs12203592 enhancer is more open and accessible in the cell lines with the CC-genotype (LP22 and DP80) than in the cell line with the CT-genotype (DP74) (Fig. 3A).

We next investigated whether the region around rs12203592 indeed acts as an enhancer element by cloning the rs12203592 region into a luciferase reporter vector. Upon transfection into human embryonic kidney cells (HEK293) no increase in luciferase expression was detected for the vectors containing the rs12203592 region with the different alleles of rs12203592 when compared with the empty vector (Fig. 3B). Whereas in G361 melanoma cell lines luciferase expression is induced for both alleles, with a statistically significantly higher activity for the rs12203592 C-allele than for the rs12203592 T-allele (Fig. 3B, $p<0.05$).

Figure 2. Identification of three regulatory elements in the *IRF4* locus; the *IRF4* promoter, the rs12203592 enhancer and a potential regulatory element in intron 7. (A) IGV genome browser (24) shows a 39 kb window of the *IRF4* locus. To investigate the chromatin for features of enhancer elements, the following tracks are included: ChIP-seq analysis in LP22 and DP74 of acetylated histone H3 (H3K27Ac), an active chromatin mark (Palstra et al, manuscript in preparation); DHS sites in epidermal skin melanocytes (20); ChIP-seq data for the transcription factor MITF in melanocytic cells (21); ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1 (22); and Phastcons conserved elements inferred from 46 way alignments of placental mammals (23). IGV genome browser shows (B) a zoomed-in frame of the region around the rs12203592 enhancer and (C) a zoomed-in frame of the region around a potential regulatory element in intron 7. The rs12203592 enhancer displays apparent characteristics of an active enhancer; strong H3K27Ac signals that are differential between LP22 and DP74, a robust DHS peak is detected and binding of the transcription factors MITF and YY1 is observed. Binding of YY1 is also observed in the potential regulatory region in intron 7, as well as a robust DHS peak, the H3K27Ac signals are modest, but differential between LP22 and DP74.

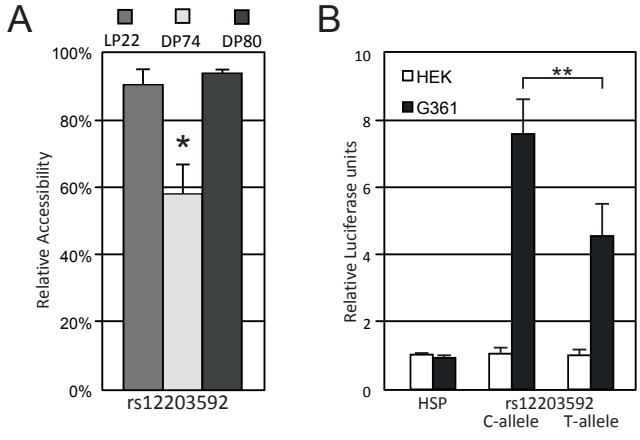


Figure 3. The region around rs12203592 displays features of an allele-dependent enhancer. (A) EpiQ analysis in the melanocyte cell lines reveals that region around rs12203592 is more open and accessible in LP22 (rs12203592-CC) and DP80 (rs12203592-CC), than in DP74 (rs12203592-CT) (* P < 0.001). (B) Luciferase reporter assay demonstrates that the activity of the rs12203592 enhancer is specific for the melanocyte lineage and differential between the rs12203592 CC-allele and the TT-allele (** P < 0.05). Data are represented as mean ± SEM.

We loaded ChIP-seq datasets for the transcription factors MITF (21) and YY1 (22) in the IGV browser (24) and we observed binding of MITF at the rs12203592 enhancer (Fig. 2B), which is consistent with previous observations (11). Additionally, a binding peak of the transcription factor YY1 is observed at the rs12203592 enhancer. When zooming out from the enhancer to an overall view of the *IRF4* locus (Fig. 2A), we observed two other YY1 peaks, one at the *IRF4* promoter and one in intron 7 of *IRF4*. Notably, both YY1 peaks coincide with robust DHS peaks and H3K27Ac-signals in the melanocytic cell lines LP22 and DP74 (Fig. 2C). Although the H3K27Ac peak in intron 7 is only moderate, it differs in amplitude between the two cell lines. Furthermore, this DHS peak in intron 7 resembles the tissue-specificity of the DHS in intron 4; it is present in melanocytes and absent in the majority of the other cell types tested (Fig. S2). However, in contrast to the DHS signal in intron 4, the DHS peak in intron 7 is not present in melanoma cells (Fig. S2). Based on the observed features of this small conserved region in intron 7 that is associated with active enhancers (18), we suggest that this region represents a potential additional regulatory element for *IRF4* in skin melanocytes.

The detection of three regions that harbor chromatin modifications specific for regulatory elements, including binding of the transcription factor YY1, demonstrate that the *IRF4* locus contains three regulatory elements; its promoter, the region around rs12203592 and an additional (potential) regulatory region at intron 7 (from here onwards referred to as the 'intron-7 YY1 element'). Moreover, the region around rs12203592 acts as a (melanocyte-specific) enhancer, and the activity of this enhancer depends on the allelic status of the rs12203592 SNP.

The rs12203592 enhancer interacts with the IRF4 promoter.

Regulatory elements, in particular enhancers, control transcription of their target gene by physically interacting with the gene promoter via chromatin loops (25), and it has been reported that multiple distal enhancers can regulate a target gene simultaneously or cooperatively (26). Furthermore, cell-type-specific transcription factors, such as MITF or YY1 not only determine the activity of the enhancer, but they are also involved in formation of the chromatin loops, either by recruiting co-factors involved in loop formation or by directly mediating loop formation (25). It has been suggested that the multifunctional transcription factor YY1 directs long-range chromatin interactions (27) and it was further demonstrated that YY1 physically interacts and co-localizes with for example cohesin and other proteins known to be involved in chromatin looping (27). We therefore reasoned that YY1 binding to the *IRF4* promoter, the rs12203592 enhancer, and the intron-7 YY1 element could mediate loop formation between these elements. In order to study this, we used 3C (28); this method enables detection of long-range chromatin interactions by trapping the interactions between chromatin segments that are in close proximity with formaldehyde. The cross-linked chromatin is subsequently digested using a restriction enzyme followed by intermolecular ligation under dilute conditions. With qPCR the relative abundance of ligation products is determined, and this is proportional to the frequency with which the different restriction fragments interact. Locus-wide 3C analysis using three different ApoI restriction fragments containing the *IRF4* promoter (P), the rs12203592 enhancer (E) and the additional YY1 element (Y), respectively, as anchor point in the melanocytic cell line LP22 (Fig. 4A-C) revealed robust interactions between the *IRF4* promoter and the rs12203592 enhancer. Interestingly, interactions between the rs12203592 enhancer and downstream located restriction fragments, which include the intron-7 YY1 element, remained high, while the interaction between the *IRF4* promoter and these downstream restriction fragments appeared to be low. Locus-wide 3C analysis in another light-pigmented melanocytic cell line (LP89; rs12203592-CC) revealed similar interaction patterns (data not shown), in agreement with the data obtained with LP22.

To gain further insight into the effect of the chromatin loops in *IRF4* we studied two melanoma cell lines BLM and G361. Characterization of BLM and G361 using HIrisPlex analysis (15) predicted light pigmentation phenotypes for both melanoma cell lines (data not shown). Furthermore, both melanoma cell lines were homozygote CC for rs12203592 (Supplementary Material, Table S1). Recently, a whole-transcriptome RNA-seq study described reduced *IRF4* expression in melanoma cells when compared with that in skin melanocytic cells (29). Indeed we detected borderline statistically significantly reduced mRNA levels of *IRF4* in the G361 cells when compared with that in the melanocytic cell lines also carrying the rs12203592 CC-genotype (p=0.056), while the expression of *IRF4* is fully absent in the BLM melanoma cell line (Fig. 1D).

We next inquired how this on/off difference of *IRF4* expression in the melanoma cell lines is reflected by the formation of chromatin loops in the *IRF4* locus. Locus-wide 3C analysis using the three different ApoI restriction fragments containing the *IRF4* promoter, the rs12203592 enhancer, and the additional YY1 element respectively, as anchor point in the melanoma cell lines G361 and BLM revealed that in G361 cells a chromatin loop is

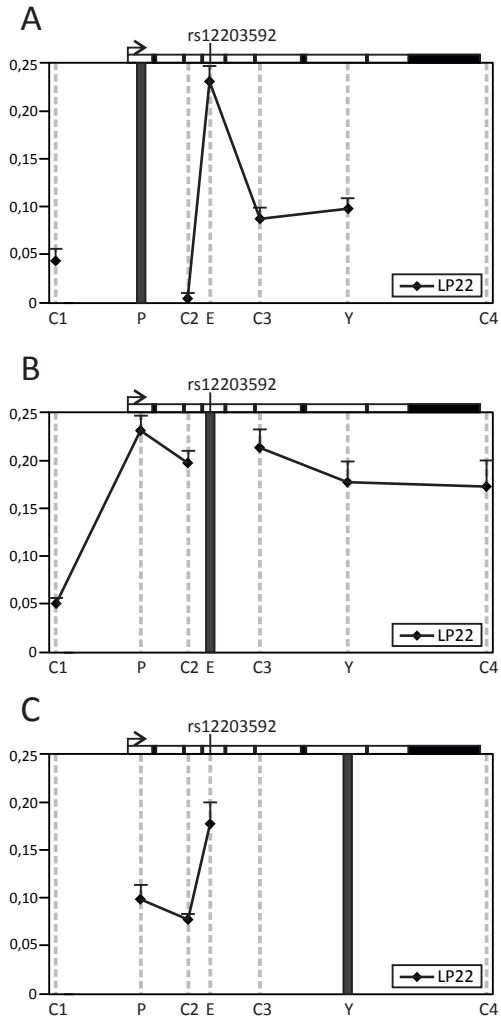


Figure 4. A long-range chromatin loop is formed between the rs12203592 enhancer and the *IRF4* promoter in skin melanocytes. Locus-wide cross-linking frequencies observed in LP22 cells (rs12203592-CC). The analyzed region of the *IRF4* locus is shown on top of each graph, the x-axis depicts the location of the fragments analyzed; C1-C4, control regions 1-4; P, *IRF4* promoter; E, rs12203592 enhancer; Y, intron-7 YY1 element. Black shading, the position of the 'fixed' restriction fragment (the anchor point), gray-shaded dotted lines, position of the other restriction fragments analyzed. Cross-linking frequencies are shown for a restriction fragment containing (A) the *IRF4* promoter, (B) the rs12203592 enhancer and (C) the intron-7 YY1 element. Robust interactions are observed between the *IRF4* promoter and the rs12203592 enhancer (A) and (B), only modest interactions are observed between the rs12203592 enhancer and the intron-7 YY1-element (B) and (C), while no interaction is detected between the *IRF4* promoter and the intron-7 YY1-element (A) and (C). Data are represented as mean ± SEM.

formed between the rs12203592 enhancer and the *IRF4* promoter, while in BLM cells this chromatin loop is absent (Fig. 5A and B), which agrees with the difference in *IRF4* expression observed between G361 and BLM cells. However, in contrast to what we observed in our melanocyte cell lines, the interaction between the rs12203592 enhancer and the intron-7 YY1 element containing downstream restriction fragments appeared to be low in both G361 and BLM cells (Fig. 5B and C), which in combination with the DNaseI Hypersensitivity data (Supplementary Material, Fig. S2) suggests that a melanocyte-specific interaction between the rs12203592 enhancer and the intron-7 YY1 element exists.

From these data we conclude that the rs12203592 enhancer regulates *IRF4* expression by directly contacting the *IRF4* promoter by formation of a chromatin loop. Furthermore, a possible interaction was detected between the rs12203592 enhancer and

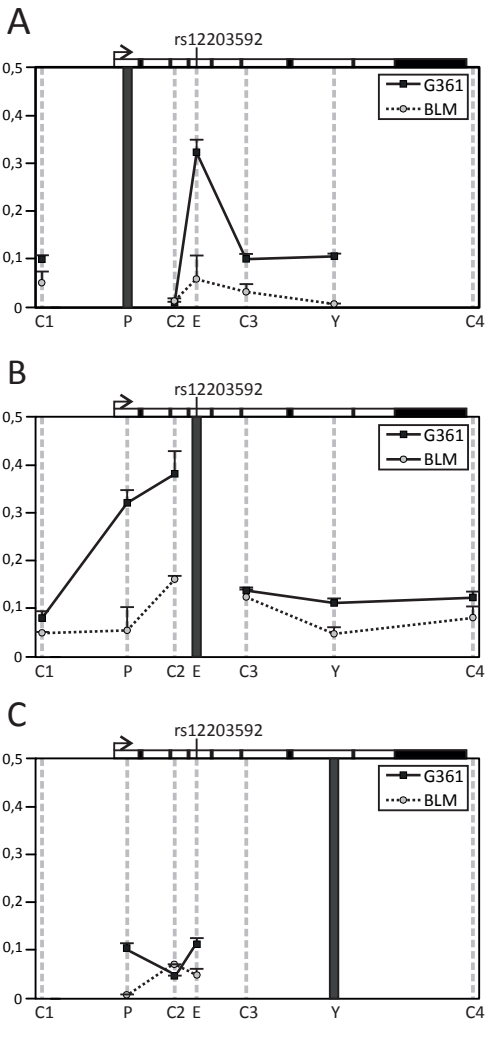


Figure 5. The interaction between the rs12203592 enhancer and the *IRF4* promoter correlates with the expression of *IRF4* in melanoma cell lines. Locus-wide cross-linking frequencies observed in G361 and BLM melanoma cells. The analyzed region of the *IRF4* locus is shown on top of each graph, the x-axis depicts the location of the fragments analyzed; C1-C4, control regions 1-4; P, *IRF4* promoter; E, rs12203592 enhancer; Y, intron-7 YY1 element. Black shading, the position of the 'fixed' restriction fragment (the anchor point), gray-shaded dotted lines, position of the other restriction fragments analyzed. Cross-linking frequencies are shown for a restriction fragment containing (A) the *IRF4* promoter, (B) the rs12203592 enhancer and (C) the intron-7 YY1 element. Interactions are observed between the *IRF4* promoter and the rs12203592 enhancer in G361 cells (A) and (B), no interactions are detected between the intron-7 YY1 element and the *IRF4* promoter or the rs12203592 enhancer in G361 cells, while in BLM cells all interactions are absent (A), (B) and (C). Data are represented as mean ± SEM.

the intron-7 YY1 element in skin melanocytes; remarkably the presence of this interaction was not detected in the melanoma cell lines possibly reflecting the absence of DHS signals in this region.

Differential IRF4 expression does not correlate with the availability of key transacting factors.

Differential availability of (tissue-specific) TFs in the different samples, rather than modulation of TF binding at the rs12203592 enhancer, could in principle also explain the differential expression of *IRF4* we and others observed. Using online-available ChIP-seq datasets for the transcription factors MITF (21) and YY1 (22) we observed binding signals for both TFs in the rs12203592 enhancer region, which confirms previous data (11). Furthermore, it

was previously demonstrated that the transcription factor TFAP2α binds to this region, and it has been suggested that the binding sequence of TFAP2α is altered by rs12203592 (4, 11). We therefore studied the expression of *TFAP2α*, *MITF* and *YY1* in relation to the pigmentation phenotypes as well as to the rs12203592 genotypes in our samples sets. In the epidermal skin samples, we observed no correlation between the expression patterns of *TFAP2α* or *MITF* and the rs12203592 genotype (Supplementary Material, Figs S3A and S4A) or the pigmentation phenotype (Supplementary Material, Figs S3B and S4B). Furthermore, we also did not detect an effect of the expression patterns of *YY1* on the transcription levels of *IRF4*, even though *YY1* expression correlated with the pigmentation phenotype of the samples (Supplementary Material, Figs S5A and B). In the melanocytic cell lines, the expression levels of *MITF* and *YY1* correlated with the pigmentation phenotypes; however, this correlation was not reflected by either the rs12203592 genotypes of the cell lines or the expression levels of *IRF4* (Supplementary Material, Figs S4C and S5C). Furthermore, the expression levels of *TFAP2α* did not correlate with either the pigmentation phenotypes or the rs12203592 genotypes of the cell lines (Supplementary Material, Fig. S3C). In the G361 cells, *MITF* expression was statistically significantly up-regulated ($p<0.005$), while *IRF4* expression was reduced when compared with the average expression of *MITF* and *IRF4* in the melanocytic cell lines with similar rs12203592 genotypes (Supplementary Material, Fig. S4C and Fig. 1D). Consistent with the previous findings (29), *MITF* expression was dramatically reduced (Supplementary Material, Fig. S4C) and *YY1* expression was strongly elevated (Supplementary Material, Fig. S5C) in BLM cells in which *IRF4* expression was fully absent, when compared with the expression of *MITF* and *YY1* ($P<0.001$ and <0.04) in melanocytic cell lines.

Based on the obtained expression analyses in the epidermal skin samples, the melanocytic cells and the melanoma cell lines, we conclude that the expression levels and subsequent availability at the enhancer region of the transcription factors *TFAP2α*, *MITF* and *YY1* do not explain the observed differences in activity of the rs12203592 enhancer and transcription levels of *IRF4* in our samples.

Allele-specific characterization of the intronic rs12203592 enhancer in IRF4

Activity of an enhancer is mediated by a broad variety of factors, both ubiquitous and cell-type specific, that modulate the specific features of the enhancer, such as modifications of histones, accessibility of the chromatin and loop formation. We have demonstrated that the availability of the three transacting factors that are presumably involved in the transcriptional regulation of *IRF4* does not correlate with the allele-dependent differences in *IRF4* expression. However, it remains possible that additional unknown factors modulate the transcriptional regulation of *IRF4* instead of the allelic differences of rs12203592. In principle, if a signal, such as H3K27Ac enrichment or chromatin accessibility, is dependent on the allelic status of an SNP, this should be detectable using an allele-specific approach in a cell-line heterozygous for this SNP. In contrast, if the signal is independent of the allelic status of a SNP the allelic balance of the DNA fragments that give rise to the signal would be equal in a heterozygote cell line for this SNP since the entire cellular environment including potentially involved *trans*-acting factors are the same for both alleles. In order to test this

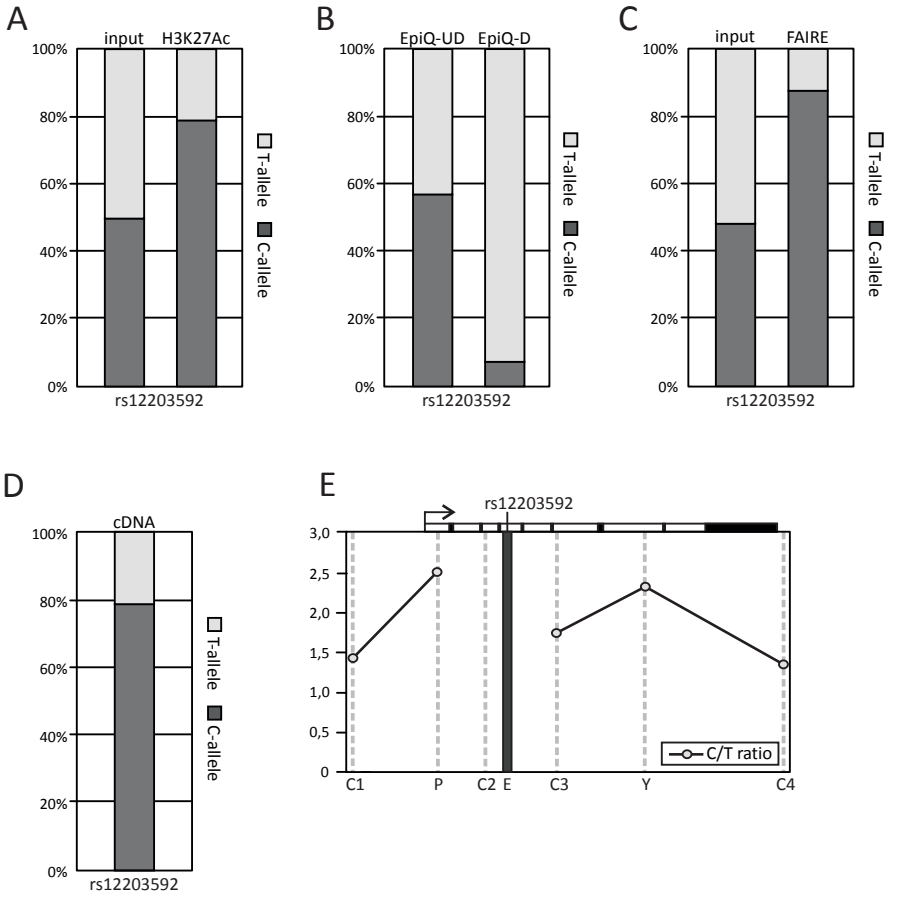


Figure 6. Characteristics of the intronic rs12203592 enhancer are strongly depending on the allelic status of rs12203592. Allele-specific analysis in the heterozygote melanocytic cell line DP74 of (A) H3K27Ac enrichment; chromatin accessibility measured with (B) EpiQ and (C) FAIRE, (D) primary transcripts and (E) loop formation detected by 3C using the rs12203592 enhancer as anchor point. Allele-specific analysis of input (H3K27Ac-ChIP and FAIRE analysis) and undigested (EpiQ) controls demonstrate equal C/T ratios of ~50/50 (A-C), while (A) the C/T ratio for the H3K27Ac signal enrichment is 79/21, (B) the C/T ratio in the EpiQ sample is 7/93, indicating that the T-allele is protected from digestion due to a closed chromatin conformation, (C) the C/T ratio of the FAIRE sample is 88/12, confirming accessible chromatin with the C-allele, and closed chromatin with the T-allele. (D) Allele-specific analysis of primary transcripts demonstrates that the majority of the transcripts enclosing the region around rs12203592 contain the C-allele. (E) The C/T ratios of all restriction fragments analyzed are higher than 1; however, increased C/T ratios are observed for the restriction fragments that contain the *IRF4* promoter (C/T: 2.5) and the intron-7 YY1 element (C/T: 2.3), indicating that the detected long-range chromatin loops between the rs12203592 enhancer and the *IRF4* promoter as well as the intron-7 YY1 element is more mediated by the C-allele than the T-allele.

hypothesis we used the heterozygote melanocytic cell line DP74 to detect allele-specific primary *IRF4* transcripts, and H3K27Ac, FAIRE and EpiQ signals at the rs12203592 enhancer.

We PCR-amplified the region around rs12203592 that was captured with the different assays and sequenced the amplicons. Subsequently, the balance between the fragments with either the rs12203592 T-allele or the rs12203592 C-allele was determined. As expected, the C/T ratio in the control samples (ChIP input; without H3K27Ac-enrichment and undigested EpiQ samples) was ~50/50 (Fig. 6A and B). Conversely, in the H3K27Ac-ChIP sample the C/T ratio was 79/21 (Fig. 6A and Supplementary Material, Fig. S6A) indicating that the C-allele is more enriched for H3K27Ac than the T-allele. In contrast, in the EpiQ sample the C/T ratio was 7/93, which implies that the T-allele is protected from digestion indicating a closed chromatin conformation around the rs12203592 enhancer (Fig. 6B and Supplementary Material, Fig. S6B). Likewise, FAIRE analysis followed by sequencing analysis to determine the allelic ratio of rs12203592 confirmed that the C-allele has an open chromatin configuration while the T-allele is inaccessible (Fig. 6C and Supplementary Material, Fig. S6C).

Since the rs12203592 enhancer is located in intron 4 of the *IRF4* gene, it is transcribed as part of the primary transcript. The C/T balance of rs12203592 in the primary transcripts in DP74 cells indicates that the majority of the RNA molecules are transcribed from the C-allele (Fig. 6D and Supplementary Material, Fig. S6D), which is in agreement with the notion that the C-allele of the rs12203592 enhancer is more active than the T-allele. We next investigated whether these allelic dis-balances are also reflected in differences in loop formation between the regulatory elements. For this, we prepared 3C samples from DP74 cells, used the rs12203592 enhancer as anchor point, and sequenced the 3C PCR-products of five distinct captured regions in the *IRF4* locus. For all restriction fragments analyzed, including the control regions (C1, C3 and C4), the C/T ratio was higher than the allelic ratio of 1. Importantly however, for the fragments that contain the *IRF4* promoter and the intron-7 YY1 element this ratio was shifted toward 2.5 (Fig. 6E and Supplementary Material, Fig. S6E). Our allele-specific 3C analysis thus demonstrates that the C-allele of rs12203592 is more actively involved in chromatin looping, and the chromatin loops are preferentially formed with the *IRF4* promoter and the intron-7 YY1 element.

Taken together, we demonstrated that the characteristics of the rs12203592 enhancer strongly depend on the allelic status of rs12203592; with the C-allele present, the region around rs12203592 is highly accessible, it is enriched for the active enhancer mark H3K27Ac and loops are formed towards the *IRF4* promoter, resulting into increased *IRF4* expression that originates from the rs12203592 C-allele. In contrast, with the T-allele present, the chromatin is closed at the region around rs12203592, it is less enriched for H3K27Ac, loop formations are less prominent and subsequently the *IRF4* expression is reduced. Our allele-specific 3C analysis also confirms our previous observation that a specific chromatin loop is formed between the rs12203592 enhancer and the intron-7 YY1 element and reveals that it is dependent on the allelic status of rs12203592.

Discussion

Enhancers are DNA elements involved in transcriptional regulation, which contain short DNA motifs that act as binding sites for sequence-specific transcription factors (25). Upon binding of specific transcription factors to the enhancer, co-activators and/or co-repressors are recruited. This combination of cues resulting from factors either directly or indirectly

interacting with the enhancer determines the activity of that enhancer (18) and as such, controls transcription of the target gene through physical contacts with the core promoter of that gene (30). The enhancer can act on a gene even if the distance between the enhancer and the promoter is relatively large. It has been shown in various studies that this physical distance is overcome by the formation of long-range chromatin loops that bring regulatory elements like enhancers and promoters in close proximity with each other (31).

Several studies showed that non-coding SNPs that were identified through GWA studies to be highly associated with a certain phenotype can be located in enhancer elements regulating transcription of a target gene, and the activities of those enhancers are modulated depending on the allelic status of the trait-associated SNPs (14, 32–35). Similarly, the SNP rs12203592 in intron 4 of *IRF4*, which had been significantly associated with human pigmentation phenotypes (5–10), was recently described to modulate an enhancer controlling expression of *IRF4* (4, 11).

In the present study, we further investigated the transcriptional regulation of *IRF4* in detail and extend on the previously published knowledge regarding the rs12203592 enhancer (4, 11). We found that transcription of *IRF4* not only correlates with the rs12203592 genotype in skin melanocytic cell lines but importantly also in a large set of epidermal skin samples. We demonstrated that the region around rs12203592 displays features of an enhancer element, and that these features strongly depend on the allelic status of rs12203592. Furthermore, we show that the rs12203592 enhancer physically interacts with the *IRF4* promoter via a long-range chromatin loop. Our data, especially from the allele-specific 3C analysis, suggests that this rs12203592 enhancer also interacts in an allele-dependent manner with an additional regulatory element in intron 7 of *IRF4* thereby folding the *IRF4* locus into an intricate structure. Based on previously published studies and the data obtained in the present study, we propose a model for the transcriptional regulation of *IRF4* in melanocytes. With the rs12203592 C-allele present, the transcription factor TFAP2α binds to the enhancer in intron 4 (11), which allows for the recruitment of MITF and YY1, formation of a chromatin loop to the *IRF4* promoter and potentially to the YY1 element in intron 7, resulting into proper and stable transcription of *IRF4*. With the rs12203592 T-allele, TFAP2α is unable to bind (11), which leads to reduced recruitment of MITF and YY1, less stable chromatin loops and consequently, decreased *IRF4* expression. Furthermore, the observed interaction differences of the rs12203592-enhancer with the *IRF4* promoter and the YY1 element in intron 7 might be due to a shift in equilibrium between the unfolded and interacting state that depends on the rs12203592 alleles (Fig. 7).

It is likely that other transcription factors are additionally involved in the mediation and stabilization of the chromatin loops and subsequent transcriptional regulation of *IRF4*. Analysis of transcription factor binding [using online available TF-binding prediction software, 'Promo' (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) (36)] predicts LEF1 binding at both regulatory regions in *IRF4*, as well as in the *IRF4* promoter region (Supplementary Material, Figs S7A-C), suggesting that LEF1 might also be involved in the loop formation and transcriptional activation of *IRF4* (Fig. 7). LEF1 has been shown to be involved in chromatin looping (37) and to interact with MITF (38), resulting into transcriptional activation of two other human pigmentation genes *DCT* (38) and *OCA2* (32).

The interaction of the rs12203592 enhancer with the potential regulatory element

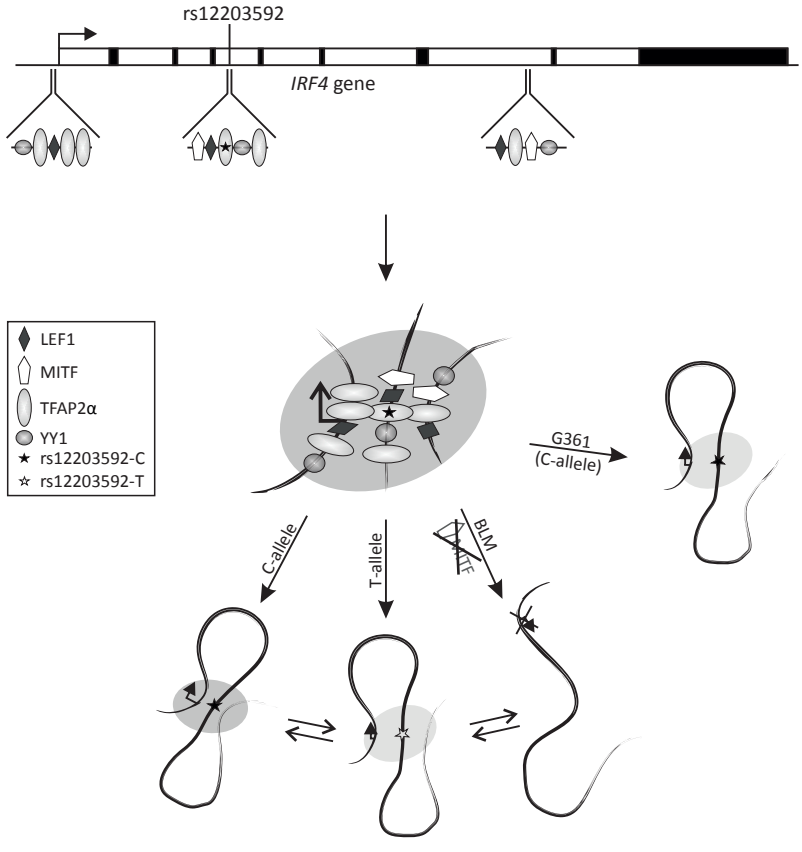


Figure 7. Proposed model for the chromatin conformation of the *IRF4* locus and subsequent transcriptional regulation of *IRF4*, depending on the allelic status of rs12203592 in melanocytes and melanoma cells. The pigmentation-associated SNP rs12203592 is located in intron 4 of *IRF4* and the region around this SNP functions as an enhancer element. The transcription factor TFAP2α acting as sequence specific DNA binding factor recognizes the rs12203592 enhancer in an allele-dependent manner, which then allows for the recruitment of the transcription factors MITF, YY1 and potential additional transcription factors like LEF1. Chromatin loops are formed between the rs12203592 enhancer and the *IRF4* promoter as well as the intron-7 YY1-element, both interactions depend on the allelic status of rs12203592. With the C-allele present, TFAP2α binds the rs12203592 enhancer, followed by recruitment of additional factors like MITF, YY1 and potentially LEF1, loop formation and proper transcriptional activation of the *IRF4* gene. The T-allele is unable to bind TFAP2α, which leads to reduced recruitment of additional factors, reduced loop formation and diminished *IRF4* expression in skin melanocytes. The different interactions between the rs12203592 enhancer and the *IRF4* promoter (as well as the intron-7 YY1 element) could be the result of a shift in equilibrium between the unfolded and interacting state that depends on the rs12203592 alleles, which is indicated by the double, opposite-directed arrows. In the G361 melanoma cell line, the TFAP2α factor presumably still binds the rs12203592 enhancer and loop formation towards the *IRF4* promoter is established, however, the loop toward the intron-7 YY1 element is disrupted, resulting in a less stable chromatin structure and consequently, diminished expression of *IRF4*. In the BLM melanoma cell line MITF is absent and no chromatin loops are formed, resulting into a linear chromatin conformation and silenced *IRF4* expression.

in intron 7 appears not to be crucial for *IRF4* activation, as we found this interaction to be almost absent in G361 melanoma cells, while *IRF4* was still transcribed. However, *IRF4* expression is reduced in this cell line when compared with that in melanocytic cell lines with the same rs12203592 genotype (CC), suggesting that this interaction is probably necessary for stabilization of the chromatin structure, resulting into proper and stable *IRF4* expression. The biological reason why this interaction with the regulatory region in intron 7 is lost in G361 cells and the exact mode-of-action of this potential enhancer element remain to be elucidated.

We recently demonstrated that it is valuable to not only focus on the associated SNP identified with a GWA study, but also to study the SNPs in linkage disequilibrium (LD) in order to discover for example the actual causal and functional variant underlying the studied associated phenotype (14). Although it appears to be straightforward in the present study that rs12203592 and not one of its LD partners is the functional SNP, we considered the possibility of an LD SNP to contribute to the transcriptional regulation of *IRF4*. Not many SNPs are in LD with *IRF4*, and of those only one, in relatively low LD (LD=0.242) is potentially interesting, as this SNP (rs3778607) is located at the border of the YY1 binding site in intron 7 of *IRF4*. It may be that this SNP modulates the activity of the additional regulatory element by altering the binding sequence of YY1, thereby contributing to the (de)stabilization of the chromatin structure and subsequent transcriptional regulation of *IRF4*, which should be investigated in future studies.

In the BLM melanoma cells we detected a linear chromatin organization in the *IRF4* locus and that *IRF4* expression was totally absent. Notably, in this cell line the master regulator of melanocyte development MITF is silenced, probably caused by and/or resulting into a complex cascade of (silencing) events, that eventually leads to a completely deregulated chromatin landscape (Fig. 7) and overall gene expression profile, which are typical features of metastatic cancer cells (39). The exact mechanism behind the loss of the *IRF4* chromatin loops, and whether the loss of these loops indeed directly results into a fully inactive gene, has to be further investigated but it seems likely that the loss of *MITF* expression in these cells plays a crucial role.

Genes often display expression patterns that are highly cell-type specific, ranging from fully silenced in one cell type to highly expressed in another. In order to achieve these differential expression patterns, transcription also needs to be regulated in a cell-type specific manner. As part of the *IRF* gene family, *IRF4* is, besides its well-established association with pigmentation phenotypes (5–10), mainly involved in immune system development and response (3) and rs12203592 was shown to be associated with (male-specific) childhood ALL (4). This study is the first to demonstrate that rs12203592 is located in an element that regulates *IRF4* expression, however with a repressive activity. With the rs12203592 C-allele present, the enhancer has a repressing effect on *IRF4* expression in three different, non-melanocytic cell lines, rather than the activating effect we and others (11) detected in skin melanocytic cells. This difference is probably due to the transcription factors involved in the activity of the rs12203592 enhancer, with TFAP2α as common transcription factor shown to be involved in all cell types tested and MITF being highly specific for skin melanocytes. Binding of this transcription factor and recruitment of other pigmentation-related transcription factors like YY1 can drive the activity of the enhancer

in an direction more favorable for transcription when compared with that in cells in which MITF is absent and other cell-type specific transcription factors, possibly with repressive functions are recruited.

The IRF4 protein encoded by the *IRF4* gene is also a transcription factor, which is thought to play a crucial role in the development of lymphoid cells. Since this gene is also expressed in melanocytic cells, it is probable that *IRF4* is also involved in pigmentation biology or melanocyte development. Indeed it was shown recently (11) that expression of *TYR*, an essential enzyme in melanogenesis, is regulated by cooperation between MITF and IRF4, which eventually leads to differences in pigmentation. As a consequence, the differences in *TYR* expression might therefore explain the strong association of rs12203592 with human skin color variation. Although it was not the scope of our study, we investigated the expression patterns of *TYR* in our sample sets and found that the expression of *TYR* indeed seems to correlate, although not statistically significantly, with the rs12203592 genotype in the epidermal skin samples (data not shown).

We did not observe a correlation between pigmentation phenotype and rs12203592 genotype in our skin biopsy sample set; however, we believe that this is most probably due to the small sample size of only 27 individuals rather than the complete absence of the correlation. The association of rs12203592 with human pigmentation phenotypes was repeatedly shown in several independent studies (6–10), that used thousands of individuals as needed for obtaining reliable outcomes in GWA studies of complex traits such as pigmentation. Additionally, a recent multivariate analysis including nine SNPs from nine genes that are strongly associated with skin color variation, demonstrated that the T-allele of rs12203592 has an significant effect on lighter skin color (F.Liu et al, manuscript in preparation). Owing to a stronger effect of other SNPs on skin color (such as rs12913832 in *HERC2* and rs4268748 in *MC1R*), it is however unlikely to detect a significant effect of rs12203592 in our small sample set consisting of only four rs12203592 CT/TT-allele samples.

Since the rise of GWA studies and the emerging data resulting therefrom, several studies have been conducted to functionally analyze the (strongly) associated DNA variants. Several of those located in non-coding regions either intergenic or intronic were demonstrated to sit in enhancer elements regulating transcription of a nearby gene. Recently, the intronic SNP rs12203592 which is highly associated with skin pigmentation was shown to be located in and modulate the activity of an enhancer controlling the gene *IRF4* (11). Here we demonstrate that this SNP not only modulates transcription factor binding and subsequent enhancer activity, but that this enhancer also physically interacts with the *IRF4* gene promoter in an allele-specific manner and with another potential regulatory element to stabilize the overall chromatin structure in order to achieve stable and cell-type specific expression of the target gene *IRF4*.

Materials and Methods

Skin sample collection and epidermal layer separation

Collection and processing of the skin samples were done as described previously (14), in short; non-sun exposed skin material that was left-over from patients undergoing plastic surgery or dermatological surgery were obtained under informed consent and patients were asked to fill in a questionnaire to obtain information concerning the pigmentation of their skin, eye and hair and more basic information as age, sex and bio-geographic ancestry. Skin color reflectance was measured at the inner side of the upper arm with spectrophotometer (Konica Minolta CM-600d), the obtained *L*, *a* and *b* values were used to calculate the individual typology angles (ITA) (40, 41) in order to categorize the skin samples according to the ITA separation of light (above 41°) and dark (under 28°) (14). This study was approved by the Medical Ethical Committee of the Erasmus MC (number MEC-2012-067).

Upon receiving the skin samples, they were cleaned and cut into small pieces, and stored in RNAlater at 4°C until further processing. The epidermal layer of the skin, including the pigment-containing melanocytic layer, was removed from the dermal layer by incubating the skin samples cut into small pieces of ± 1 mm wide and 7 mm long for 30 min at room temperature in Ammonium Thiocyanate solution (3.8% ATC in PBS) (42). The epidermal layers were then peeled off carefully and stored in RNAlater at 4°C until RNA/DNA co-isolation, or at -80°C for long-term storage.

Cell culture

HEMn-LP22 (C-0025C; lot # 200708522; Cascade Biologics, Invitrogen), HEMn-DP74 (C-2025C, lot # 6C0474; Cascade Biologics, Invitrogen) and HEMn-DP80 (C-2025C, lot # 200707980 Cascade Biologics, Invitrogen) were grown in Medium 254 supplemented with HMGS according to the manufacturer's instruction (Cascade Biologics, Invitrogen). G361, HEK293 and BLM cells were cultured in DMEM/10% FCS at 37°C/5% CO_2.

RNA/DNA co-isolation from skin samples and from cultured cells

Epidermal skin tissue was lysed and RNA and DNA were co-isolated with the Qiagen Allprep mini kit according to the manufacturer's instructions for fibrous tissue. From the different cell lines, total cellular RNA and genomic DNA were isolated with TriPure Isolation Reagent according to the manufacturer's protocol (Roche Diagnostics). DNA and RNA samples were column purified to remove PCR inhibiting substances like melanin (*OneStep*™ PCR Inhibitor Removal Kit, Zymo Research Corporation) and subsequent DNaseI digestion of the RNA samples was performed using Ambion's Turbo DNA-free kit (Applied Biosystems) following the instructions of the manufacturers.

Transcription analysis

The RT reaction was performed using RevertAid™ H Minus First Strand cDNA Synthesis Kit

(Fermentas GmbH) according to the manufacturer's instructions. Quantitative real-time PCR reactions for gene-expression analysis were performed using the iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories) with the following PCR parameters: initial denaturation at 95°C for 5 min, followed by 45 cycles at 95°C for 5 s and 60°C for 30 s, followed by a melting curve analysis. Bio-Rad Software v1.5 in combination with Excel was used to analyze the qPCR data. The reference gene *ACTB* was used to normalize the amplification signal between samples, differences in treatment and amount of input cDNA. Primers are available on request. Student's two-tailed t-test was used to determine statistical significance.

SNP genotyping

The HIrisPlex assay and the interactive HIrisPlex prediction tool were used as described by (15) to predict hair and eye color for the cell line samples, to confirm the obtained eye and hair color phenotype information of all skin biopsy samples and additionally to obtain genotypes of the SNP rs12203592.

Chromatin immuno-precipitation sequencing

ChIP was performed as described in the Millipore protocol (http://www.millipore.com/userguides/tech1/mcproto407), except that samples were cross-linked with 2% formaldehyde for 10 minutes at room temperature. To remove melanin, DNA was column purified prior to PCR (*OneStep*™ PCR Inhibitor Removal Kit, Zymo Research). ChIP-seq analysis was performed as described previously (43). Antibody used: acetylated Histone H3 K27 (#4729ab) from Abcam.

EpiQ analysis

Q-PCR–based EpiQ chromatin analysis assay (Bio-Rad Laboratories) was performed according to the manufacturer's protocol. To remove melanin, DNA was column purified prior to PCR (*OneStep*™ PCR Inhibitor Removal Kit, Zymo Research). Quantitative real-time PCR was performed using Universal SYBR green master mix (Bio-Rad Laboratories) under the following cycling conditions: 95°C for 5 min, 45 cycles of 10 s at 95°C, 30 s at 60°C and followed by a melting curve analysis. PCR primers are available on request.

Luciferase assays

Custom500bp gBlocks Gene Fragments (Integrated DNA Technologies) containing either the C- or T-allele of the rs12203592 enhancer region and XhoI and BglII restriction sites were cloned into the XhoI/BglII sites of a modified pGL3-promoter vector (in which the SV40 promoter and the SV40 3'UTR were replaced by an HSP promoter and an HSP 3'UTR) (Promega). Inserts in each construct were verified by sequencing (Baseclear). Constructs were transfected into HEK293 or G361 melanoma cells using Lipofectamine LTX (Invitrogen), and luciferase expression was normalized to Renilla luciferase expression. Primer sequences are available on request. Data represent at least five independent experiments. Student's

two-tailed *t* test was used to determine statistical significance.

Chromosome conformation capture analysis

3C analysis was performed essentially as described (44, 45) using ApoI as the restriction enzyme. To remove melanin, DNA was column purified prior to PCR (OneStep™ PCR Inhibitor Removal Kit, Zymo research). Quantitative real-time PCR (CFX96TM Real Time System, BioRad) was performed using iTaq SYBR Green Supermix with ROX (BioRad), under the following cycling conditions: 50°C for 2 min, 95°C for 10 min, 45 cycles of 15 s at 95°C, 1 min at 60°C, followed by a melting curve analysis. A random template was generated from a 975 bp custom designed gBlocks Gene Fragments (Integrated DNA Technologies). The gBlock consisted of a tandem array of ~100bp amplicons covering the ApoI restriction sites of interest within the *IRF4* locus (control regions C1-4, the *IRF4* promoter, the rs12203592 enhancer and the intron-7 YY1 element) separated by 4bp spacers. The gBlock was digested with ApoI, re-ligated and mixed with genomic DNA to serve as a control template to correct for differences in primer efficiency. PCR primers are available on request.

Allele-specific analysis

For allele-specific analysis respective templates (cDNA, gDNA and DNA enriched in FAIRE, EpiQ and ChIP experiments) were amplified with primers covering the rs12203592 SNP using the Expand Long Template PCR system (Roche). Amplicons were gel purified and sequenced (Baseclear). Allelic ratios were calculated from the peak plot files. PCR primers are available on request.

Acknowledgments

We are very thankful to all volunteers who agreed their materials to be used for this study. We acknowledge the staff, especially Jan Maerten Smit and Han van Neck, from the Department of Plastic Surgery, and Tamar Nijsten, Leonie Jacobs and Emmilia Dowlatshahi from the Department of Dermatology, both of Erasmus MC, who helped with collecting epidermal skin samples. We additionally thank the staff of the Erasmus Center of Biomics, especially Wilfred van IJcken and Rutger Brouwer, for sequencing and data-processing of the ChIP-seq samples. Lakshmi Chaitanya and Arwin Ralf from the Department of Forensic Molecular Biology of Erasmus MC are acknowledged for help with genotyping.
Funding: This study was supported by the Erasmus MC, and in part by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

References

1. Edwards, S.L., Beesley, J., French, J.D. and Dunning, A.M. (2013) Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.*, **93**, 779–797.
2. Paun, A. and Pitha, P.M. (2007) The IRF family, revisited. *Biochimie*, **89**, 744–753.
3. Gualco, G., Weiss, L.M. and Bacchi, C.E. (2010) MUM1/IRF4: A review. *Appl. Immunohistochem. Mol. Morphol.*, **18**, 301–310.
4. Do, T.N., Ucisik-Akkaya, E., Davis, C.F., Morrison, B.A. and Dorak, M.T. (2010) An intronic polymorphism of IRF4 gene influences gene transcription in vitro and shows a risk association with childhood acute lymphoblastic leukemia in males. *Biochim. Biophys. Acta*, **1802**, 292–300.
5. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., *et al.* (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.*, **39**, 1443–1452.
6. Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z., *et al.* (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.*, **4**, e1000074.
7. Liu, F., Wollstein, A., Hysi, P.G., Ankra-Badu, G. a, Spector, T.D., Park, D., Zhu, G., Larsson, M., Duffy, D.L., Montgomery, G.W., *et al.* (2010) Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet.*, **6**, e1000934.
8. Eriksson, N., Macpherson, J.M., Tung, J.Y., Hon, L.S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I. and Mountain, J. (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.*, **6**, e1000993.
9. Jacobs, L.C., Wollstein, A., Lao, O., Hofman, A., Klaver, C.C., Uitterlinden, A.G., Nijsten, T., Kayser, M. and Liu, F. (2013) Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Hum. Genet.*, **132**, 147–158.
10. Nan, H., Kraft, P., Qureshi, A. a, Guo, Q., Chen, C., Hankinson, S.E., Hu, F.B., Thomas, G., Hoover, R.N., Chanock, S., *et al.* (2009) Genome-wide association study of tanning phenotype in a population of European ancestry. *J. Invest. Dermatol.*, **129**, 2250–2257.
11. Praetorius, C., Grill, C., Stacey, S.N., Metcalf, A.M., Gorkin, D.U., Robinson, K.C., Van Otterloo, E., Kim, R.S.Q., Bergsteinsdottir, K., Ogmundsdottir, M.H., *et al.* (2013) A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell*, **155**, 1022–1033.
12. Tellez, C., McCarty, M., Ruiz, M. and Bar-Eli, M. (2003) Loss of activator protein-2alpha results in overexpression of protease-activated receptor-1 and correlates with the malignant phenotype of human melanoma. *J. Biol. Chem.*, **278**, 46632–46642.
13. Ruiz, M., Pettaway, C., Song, R., Stoeltzing, O., Ellis, L. and Bar-eli, M. (2004) Activator Protein 2alpha Inhibits Tumorigenicity and Represses Vascular Endothelial Growth Factor Transcription in Prostate Cancer Cells. *Cancer Res.*, **64**, 631-638.
14. Visser, M., Palstra, R.-J. and Kayser, M. (2014) Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. *Hum. Mol. Genet.*, **23**, 5750–5762.
15. Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W. and Kayser, M. (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.*, **7**, 98–115.
16. Walsh, S., Liu, F., Ballantyne, K.N., van Oven, M., Lao, O. and Kayser, M. (2011) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci. Int. Genet.*, **5**, 170–180.
17. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
18. Shlyueva, D., Stampfel, G. and Stark, A. (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.*, **15**, 272–286.
19. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 21931–21936.
20. Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C.,

Maddren, M., Fang, R., Heitner, S.G., *et al.* (2013) ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.*, **41**, D56-D63.
21. Strub, T., Giuliano, S., Ye, T., Bonet, C., Keime, C., Kobi, D., Le Gras, S., Cormont, M., Ballotti, R., Bertolotto, C., *et al.* (2011) Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene*, **30**, 2319–2332.
22. Li, J., Song, J.S., Bell, R.J. a, Tran, T.-N.T., Haq, R., Liu, H., Love, K.T., Langer, R., Anderson, D.G., Larue, L., *et al.* (2012) YY1 regulates melanocyte development and function by cooperating with MITF. *PLoS Genet.*, **8**, e1002688.
23. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
24. Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.
25. Marsman, J. and Horsfield, J.A. (2012) Long distance relationships: enhancer-promoter communication and dynamic gene transcription. *Biochim. Biophys. Acta*, **1819**, 1217–1227.
26. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. and Zhao, K. (2012) Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.*, **22**, 490–503.
27. Atchison, M.L. (2014) Function of YY1 in Long-Distance DNA Interactions. *Front. Immunol.*, **5**, 45.
28. De Laat, W. and Dekker, J. (2012) 3C-based technologies to study the shape of the genome. *Methods*, **58**, 189–191.
29. Zhao, H., Li, Y., Wang, S., Yang, Y., Wang, J., Ruan, X., Yang, Y., Cai, K., Zhang, B., Cui, P., *et al.* (2014) Whole transcriptome RNA-seq analysis: tumorigenesis and metastasis of melanoma. *Gene*, **548**, 234–243.
30. Lee, T.I. and Young, R.A (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
31. Palstra, R.-J. (2009) Close encounters of the 3C kind: long-range chromatin interactions and transcriptional regulation. *Brief. Funct. Genomic. Proteomic.*, **8**, 297–309.
32. Visser, M., Kayser, M. and Palstra, R.-J. (2012) HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.*, **22**, 446–455.
33. Pomerantz, M.M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M.P., Doddapaneni, H., Beckwith, C.A., Chan, J.A., Hills, A., Davis, M., *et al.* (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.*, **41**, 882–884.
34. Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C. and Nobrega, M.A. (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.*, **21**, 3255–3263.
35. Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.-D., Topol, E.J., Rosenfeld, M.G., *et al.* (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature*, **470**, 264–268.
36. Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J. and Albà, M.M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*, **18**, 333–334.
37. Yun, K., So, J.-S., Jash, A. and Im, S.-H. (2009) Lymphoid enhancer binding factor 1 regulates transcription through gene looping. *J. Immunol.*, **183**, 5129–5137.
38. Yasumoto, K., Takeda, K., Saito, H. and Watanabe, K. (2002) Microphthalmia-associated transcription factor interacts with LEF-1 , a mediator of Wnt signaling. *EMBO J.*, **21**, 2703–2714.
39. Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell*, **60**, 319–326.
40. Chardon, A., Cretois, I. and Hourseau, C. (1991) Skin colour typology and suntanning pathways. *Int. J. Cosmet. Sci.*, **208**, 191–208.
41. Del Bino, S., Sok, J., Bessac, E. and Bernerd, F. (2006) Relationship between skin response to ultraviolet exposure and skin color type. *Pigment Cell Res.*, **19**, 606–614.
42. Trost, A., Bauer, J.W., Lanschützer, C., Laimer, M., Emberger, M., Hintner, H. and Onder, K. (2007) Rapid, high-quality and epidermal-specific isolation of RNA from human skin. *Exp. Dermatol.*, **16**, 185–190.
43. Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.-J., Stevens, M., Kockx, C., van Ijcken, W., *et al.* (2010) The genome-wide dynamics of the binding of

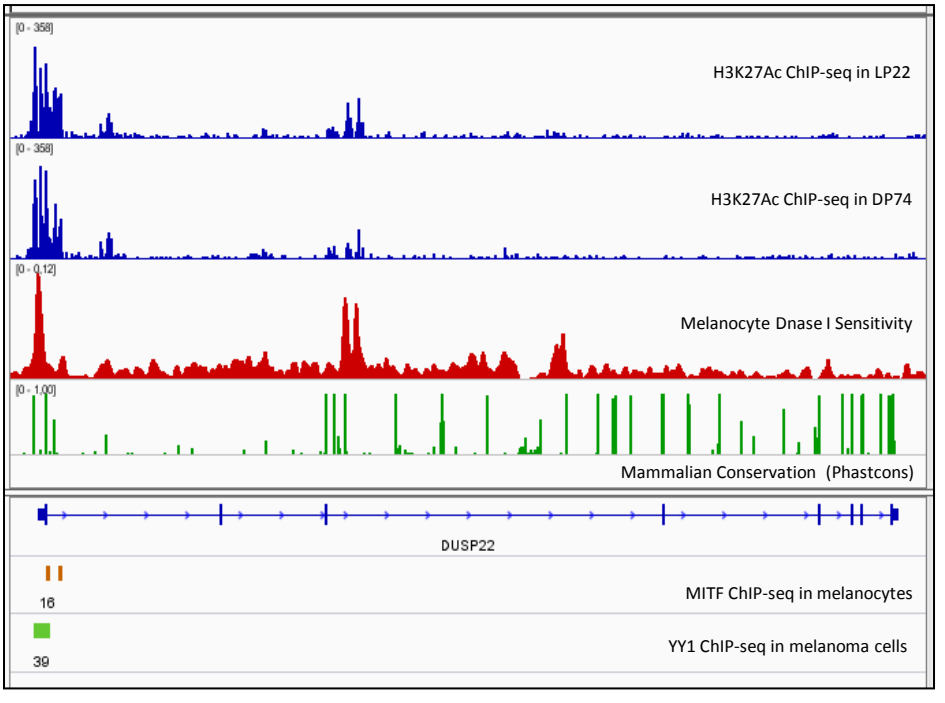
Ldb1 complexes during erythroid differentiation. *Genes Dev.*, **24**, 277–289.
44. Palstra, R-J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F. and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat. Genet., 35, 190–194.
45. Hagège H, Klous P, Braem C, Splinter, E., Dekker, J., Cathala, G., de Laat, W. and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protoc., 2, 1722–1733.

Supplemental tables and figures

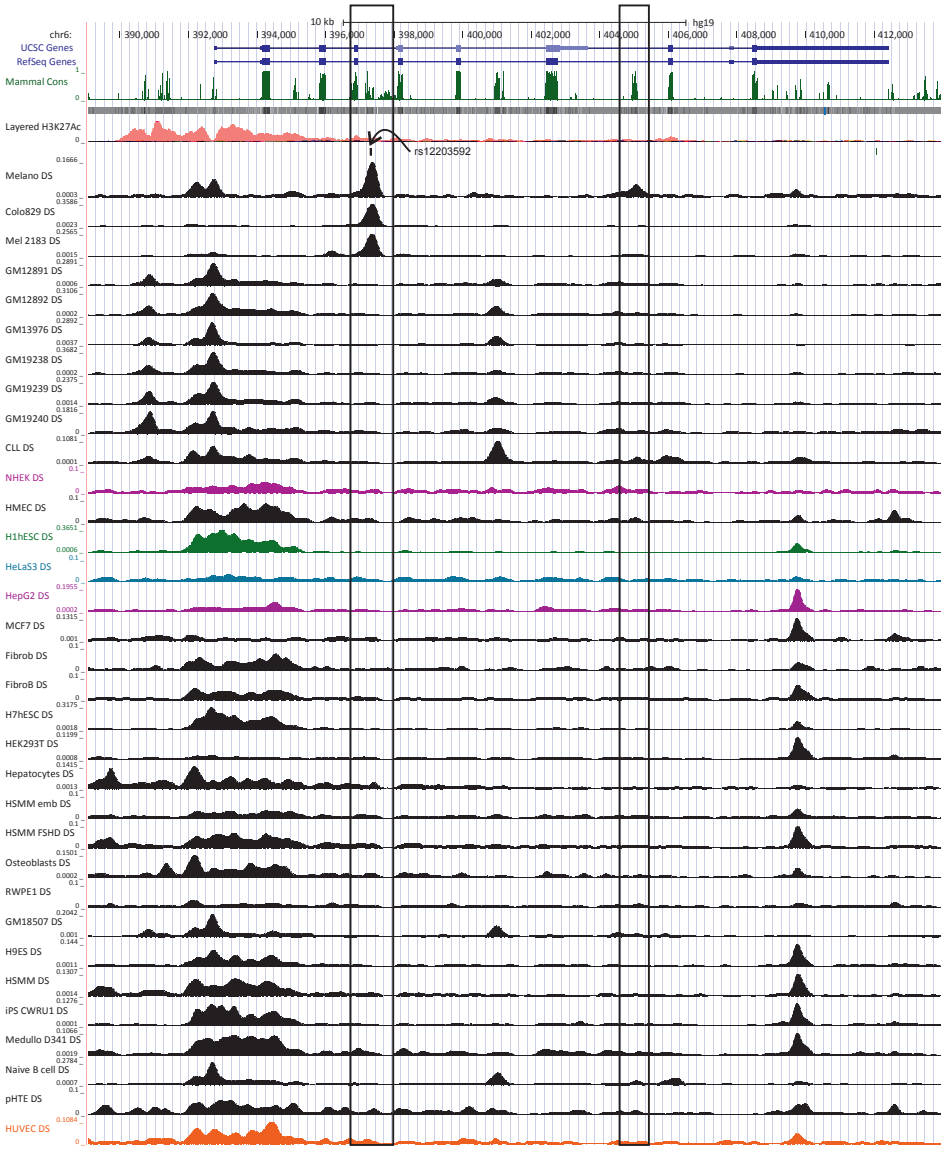
Supplemental table S1.

sample number	geographical ancestry	pigmentation category	rs12203592 genotype
Epidermal Skin Samples			
1	Surinam	dark	CC
2	Netherlands	light	CC
3	Surinam	dark	CC
4	Netherlands	light	**CT**
5	Netherlands	light	**CT**
6	Netherlands	light	CC
7	Curacao	dark	CC
8	Netherlands	light	**TT**
9	Netherlands	light	CC
10	Netherlands	light	CC
11	Netherlands	light	CC
12	Surinam	dark	CC
13	Netherlands	light	CC
14	Surinam	dark	CC
15	Netherlands	light	CC
16	Netherlands	light	**CT**
17	Netherlands	light	CC
18	Netherlands	light	CC
19	Netherlands	light	CC
20	Eritrea	dark	CC
21	Netherlands	light	CC
22	Surinam	dark	CC
23	Surinam	dark	CC
24	Surinam	dark	CC
25	Curacao	dark	CC
26	Surinam	dark	CC
27	Curacao	dark	CC
28	Netherlands	light	CC
29	Netherlands	light	CC
Skin melanocytes			
LP22	European	light	CC
DP74	African	dark	**CT**
DP80	African	dark	CC
melanoma cell line			
BLM	Caucasian	light	CC
G361	Caucasian	light	CC

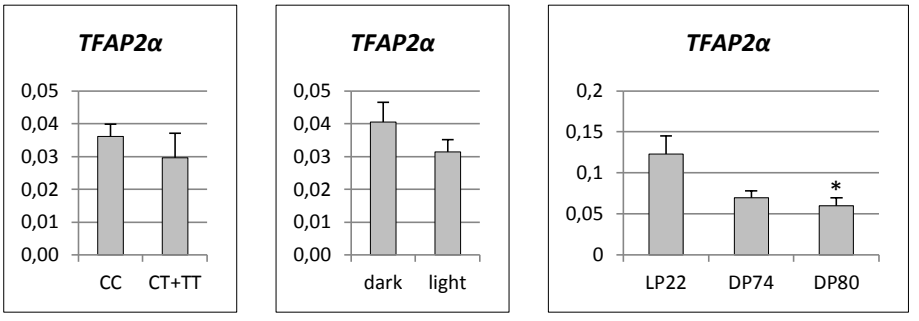
Geographical ancestry information, phenotype and rs12203593 genotype of skin epidermal samples, melanocyte cells and melanoma cell lines. Geographical ancestry information is based on place of birth of the sample donor and that of their parents and grandparents. Skin color is categorized according to the ITA values as is described previously (14) and in the materials and methods section. Genotypes of rs12203592 are obtained with the HIrisPlex assay (15).



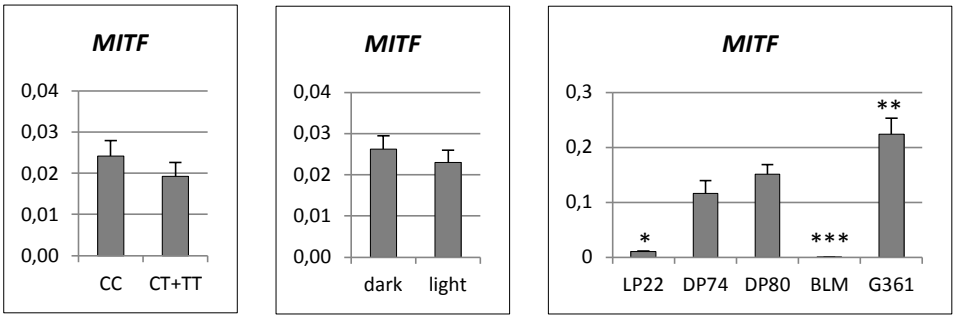
Supplemental figure S1. H3K27Ac enrichment is not differential between LP22 and DP74 in a non-pigmentation related region. IGV genome browser (24) shows a 63 kb window of the *DUSP22* locus. The following tracks are included: ChIP-seq analysis in LP22 and DP74 of acetylated histone H3 (H3K27Ac), an active chromatin mark (Palstra et al, manuscript in preparation); DNaseI hypersensitive (DHS) sites in epidermal skin melanocytes (20); ChIP-seq data for the transcription factor MITF in melanocytic cells (21); ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1 (22); and Phastcons conserved elements inferred from 46 way alignments of placental mammals (23). No differences in H3K27Ac enrichment are observed at the promoter region of *DUSP22* between LP22 and DP74.



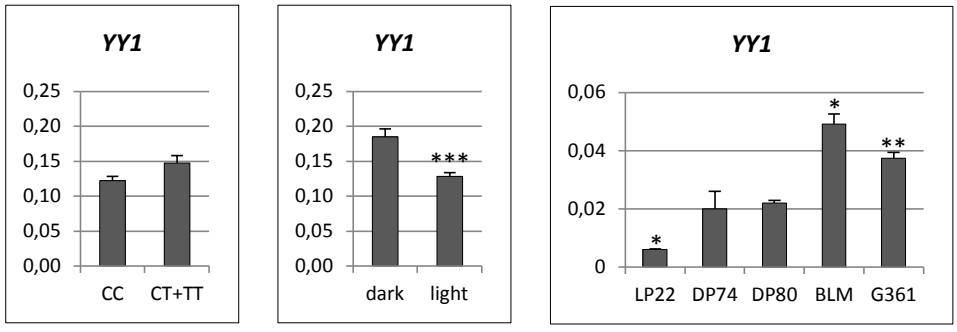
Supplemental figure S2. Tissue-specificity of the rs12203592 enhancer and the intron-7 YY1-element. Tracks from the UCSC browser displaying DNaseI HyperSensitivity (DHS) in the *IRF4* locus in a variety of cell types and tissues. The regions of the rs12203592 enhancer and the intron-7 YY1-element are boxed. DHS signals for the rs12203592 enhancer are only present in melanocytes (top DHS track) for both elements, as well as in two melanoma cell lines (DHS track 2 and 3 from the top).



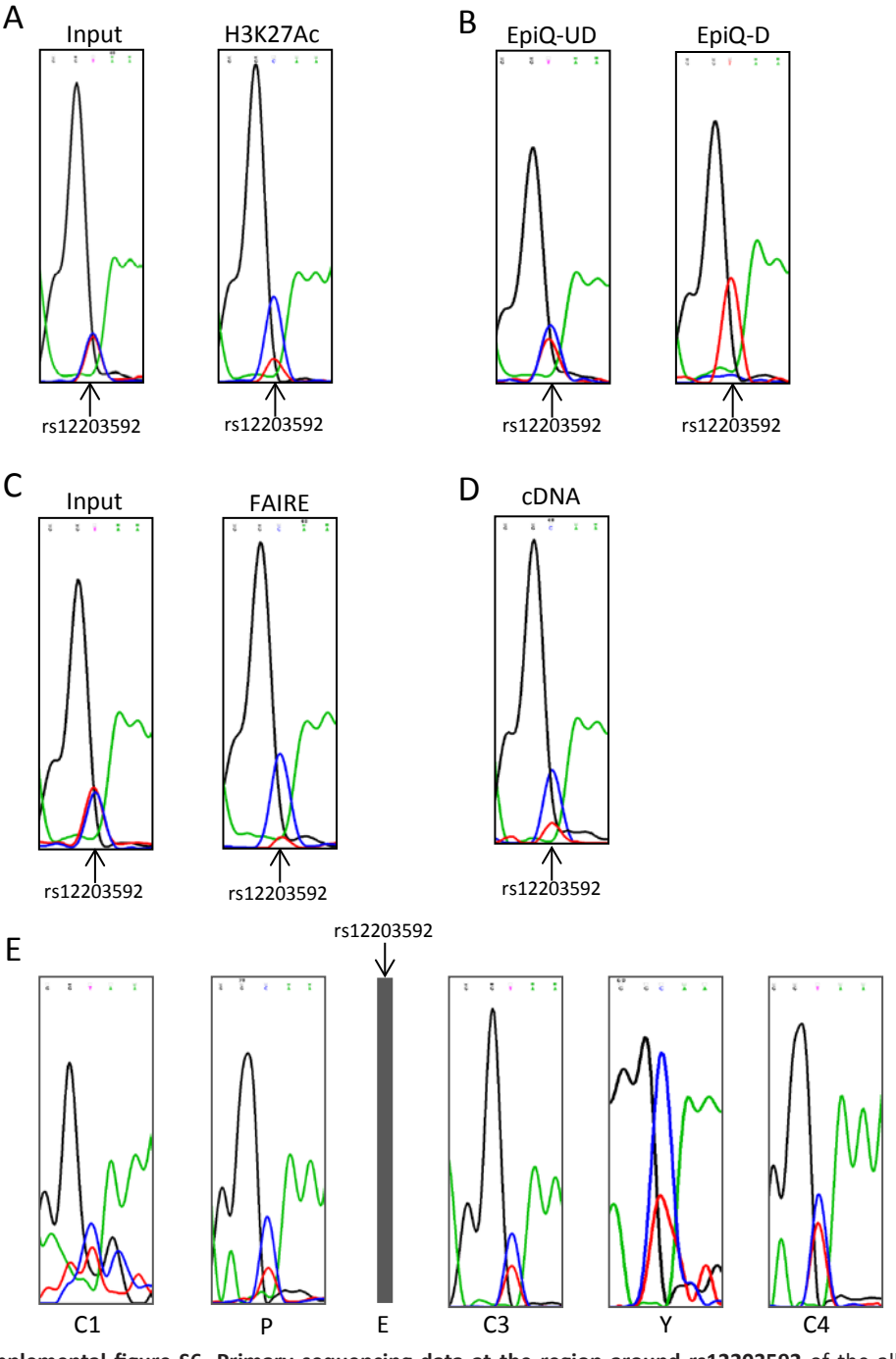
Supplemental figure S3. Transcription of *TFAP2α* in epidermal skin samples and melanocyte cell lines. (A) RT-qPCR analysis of *TFAP2α* transcripts in skin epidermal samples with either the rs12203592 CC-genotype (n=25) or with the combined rs12203592 CT- and TT-genotypes (N=4) demonstrates that *TFAP2α* is not differentially expressed between CC and CT/TT genotypes. (C) RT-qPCR analysis of *TFAP2α* transcripts in skin epidermal samples with either dark (n=12) or light (n=17) phenotype indicates no differential expression of *TFAP2α* between the two phenotype categories. (D) RT-qPCR analysis of *TFAP2α* transcripts in three melanocyte cell lines demonstrates that expression of *TFAP2α* is lower in the dark melanocyte cell lines (DP74 and DP80) than in the light melanocyte cell line (LP22). Each individual gene-expression analysis is performed in triplicate and normalized to an endogenous control (*ACTB*). Averaged expression values for the genotype (CC; CT+TT) and skin color phenotype (light; dark) categories of the skin epidermal samples are calculated after normalization with *ACTB*. Data are represented as mean ± SEM; * P < 0.05.



Supplemental figure S4. Transcription of *MITF* in epidermal skin samples, melanocyte cell lines and melanoma cell lines. (A) RT-qPCR analysis of *MITF* transcripts in skin epidermal samples with either the rs12203592 CC-genotype (n=25) or with the combined rs12203592 CT- and TT-genotypes (N=4) demonstrates that *MITF* is not differentially expressed between CC and CT/TT. (C) RT-qPCR analysis of *MITF* transcripts in skin epidermal samples with either dark (n=12) or light (n=17) phenotype indicates no differential expression of *MITF* between the two phenotype categories. (D) RT-qPCR analysis of *MITF* transcripts in three melanocyte cell lines and two melanoma cell lines. This reveals that expression of *MITF* is significantly lower in the light-pigmented melanocyte cell line (LP22) than in the dark-pigmented melanocyte cell lines (DP74 and DP80). Expression of *MITF* is dramatically reduced in BLM melanoma cells, whereas in G361 melanoma cells *MITF* is up-regulated as compared to the expression of *MITF* in the melanocyte cell lines. Each individual gene-expression analysis is performed in triplicate and normalized to an endogenous control (*ACTB*). Averaged expression values for the genotype (CC; CT+TT) and skin color phenotype (light; dark) categories of the skin epidermal samples are calculated after normalization with *ACTB*. Data are represented as mean ± SEM; * P < 0.005; ** P < 0.0001; *** P < 0.00001.



Supplemental figure S5. Transcription of *YY1* in epidermal skin samples, melanocyte cell lines and melanoma cell lines. (A) RT-qPCR analysis of *YY1* transcripts in skin epidermal samples with either the rs12203592 CC-genotype (n=25) or with the combined rs12203592 CT- and TT-genotypes (N=4) demonstrates that *YY1* is not differentially expressed between CC and CT/TT. (C) RT-qPCR analysis of *YY1* transcripts in skin epidermal samples with either dark (n=12) or light (n=17) phenotype indicates that *YY1* is significantly differentially expressed between the two phenotype categories. (D) RT-qPCR analysis of *YY1* transcripts in three melanocyte cell lines and two melanoma cell lines. This reveals that expression of *YY1* is significantly lower in the light-pigmented melanocyte cell line (LP22) than in the dark-pigmented melanocyte cell lines (DP74 and DP80). Expression of *YY1* is up-regulated in both melanoma cell lines as compared to the expression of *YY1* in the melanocyte cell lines. Each individual gene-expression analysis is performed in triplicate and normalized to an endogenous control (*ACTB*). Averaged expression values for the genotype (CC; CT+TT) and skin color phenotype (light; dark) categories of the skin epidermal samples are calculated after normalization with *ACTB*. Data are represented as mean ± SEM; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.00001$.



Supplemental figure S6. Primary sequencing data at the region around rs12203592 of the allele-specific analyses in the heterozygote melanocytic cell line DP74 of (A) H3K27Ac enrichment; chromatin accessibility measured with (B) EpiQ and (C) FAIRE; (D) primary transcripts and (E) loop formation detected by 3C using the rs12203592 enhancer as anchor point.

TFAP2a/MITF/YY1/ LEF1

(A) Promoter region

GTTCTCTTTGGACCATTCCTCCGTCTTCCGTTACACGCTCTGCAAAGCGAAGTCCCCTTCGCACCAGATTCCCGCTACTACAC

GCCCCCCATTTCCCGCCCTGGCCACATCGCTGCAGTTTAGTGATTGACTGGCCTCCTGAGGTCCTGGCGCAAAGGCGAGAT

TCGCATTTCGCACCTCGCCCTTCGCGGGAAACGGCCCCAGTGACAGTCCCCGAAGCGGCGCGCGCCCGGCTGGAGGTGCG

CTCTCCGGGCGCGGCGCGCGGAGGGTCGCCAAGGGCGCGGGAACCCCACCCCGGCCGCGGCAGCCCCCAGCCTTCACGC

CGGCCCTGAGGCTCGCCCGCCCGGCCGGCCCCGGCTCTCGGCTTGCAAAGTCCCTCTCCCCAGTCCAACCCCCGGCC

(B) rs12203592 enhancer region

CAGCCCAGGTATGGTGGAGGGCACTGGGCTCCCTGAGGGCGAGGCTGTGTGGGCCAGCTGCCCACATGGCCAGAGAACC

ACAGCAGCCCAGACAGCAGAACTTGCCATTTGCTATGGCTGCTCCAACAGCCCAGAAAAACCCCAGGTCACTGAACGAAT

GTCTCACTTTCCACACGGTGCTGCCATTGGTGTGGATTTTAAGTTGGGGAGGGTCGGGCGTGTCCGCCTGTTGGAATATGC

TTCTCAGGTCTTCTGGGAAACAGATGTTTTGTGGAAGTGGAAGATTTTGGAAGTAGTGCCTTATCATGTGAAACCACAGGG

CAGCTGATCTCTTCAGGCTTTCTTGATGTGAATGACAGCTTTGTTTCATCCACTTTGGTGGGTAAAAGAAGGCAAATTCCCC

TGTGGTACTTTTGGTGCCAGGTTTAGCCATATGACGAAGCTTTACATAAAACAGTACAAGTATCTCCATTGTCCTTTATGAT

CCTCCATGAGTGTTTTCACTTAGTCTGATGAAGGGTTCACTCCAGTCTTTTCGGATGATAAAATGCTTCGGCTGTCAGTCTA

(C) Intron-7 YY1-element region

GGTCTTGGCTTTCATTCCTTCCTCTGTAACCAAGGGCTGTGCTCTTTGCCCACTGCAGCCTCTCACCTCAAAGTGTTCACTGA

GGTTCAAGGAGGATGCTGTGAACGTAAAAACTGCAGCTGTGCCAACCAGCTTCTGCATAATTAAGGATTCCCACCAAACA

CTCTCATGTTATCTAGGGTTGGAGCCATGCTTTCTCAGAGAATTGTGCCAACTCGCCATTCTGATTAGCCTGTGTAGGTGTA

GTCTCAGATCACGGCAGTGTGAATGTATTTTACAGATTCTGACTAAGTCATTTGG

Supplemental figure S7. Prediction of transcription factors in three regulatory regions in the *IRF4* locus. Potential binding sites for the transcription factors TFAP2α, MITF, YY1 and LEF1 in the (A) *IRF4* promoter, (B) the rs12203592 enhancer and (C) the intron-7 YY1-element, are predicted using the online available tool 'Promo' (36) (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3). TFAP2α binding is predicted at several positions in all three regulatory elements (A-C), the SNP rs12203592 is located in one of those predicted TFAP2α binding sites (bolded **C/T** in (B)), which confirms previous notions (4, 11)a transcription factor with no known role in melanocyte biology, is strongly associated with sensitivity of skin to sun exposure, freckles, blue eyes, and brown hair color. Here, we demonstrate that this SNP lies within an enhancer of IRF4 transcription in melanocytes. The allele associated with this pigmentation phenotype impairs binding of the TFAP2A transcription factor that, together with the melanocyte master regulator MITF, regulates activity of the enhancer. Assays in zebrafish and mice reveal that IRF4 cooperates with MITF to activate expression of Tyrosinase (TYR. Binding of MITF is predicted in the rs12203592 enhancer (B) and in the intron-7 YY1 element (C), but not in the *IRF4* promoter (C), while binding of YY1 and LEF1 is predicted in this promoter region (A) as well as in the other two regulatory regions (B) and (C).

Chapter 5

Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up

Fan Liu*, Mijke Visser*, David L. Duffy, Pirro G. Hysi, Leonie C. Jacobs, Oscar Lao, Kaiyin Zhong, Susan Walsh, Lakshmi Chaitanya, Andreas Wollstein, Gu Zhu, Grant W. Montgomery, Anjali K. Henders, Massimo Mangino, Daniel Glass, Veronique Bataille, Richard A. Sturm, Fernando Rivadeneira, Albert Hofman, Wilfred F.J. van IJcken, André G. Uitterlinden, Robert-Jan T.S. Palstra, Timothy D. Spector, Nicholas G. Martin, Tamar E.C. Nijsten, and Manfred Kayser, for the International Visible Trait Genetics (VisiGen) Consortium

*Equal contribution

Submitted

Abstract

In the International Visible Trait Genetics (VisiGen) Consortium, we investigated the genetics of human skin color by combining a series of genome-wide association studies (GWAS) in a total of 17,262 Europeans with functional follow-up of discovered loci. Our GWAS provide the first genome-wide significant evidence for chromosome 20q11.22 being explicitly associated with skin color in Europeans. In addition, genomic loci at 5p13.2, 6p25.3, 15q13.1, and 16q24.3 were confirmed to be involved in skin coloration. In follow-up gene expression and regulation studies of 22 genes in 20q11.22, we highlighted two novel genes *EIF2S2* and *GSS*, serving as competing functional candidates in this region. A genetically inferred skin color score obtained from the 9 top-associated SNPs from 9 genes in 940 worldwide samples (HGDP-CEPH) showed a clear gradual pattern in Western Eurasians similar to the distribution of physical skin color, suggesting the used 9 SNPs as suitable markers for DNA prediction of skin color, relevant in future forensic and anthropological investigations.

Introduction

Whilst the principal genes influencing eye and hair color are now largely identified, current knowledge on the genetic basis of skin color variation is still limited (Liu et al. 2013). A better understanding of human skin color genetics is highly relevant for medicine i.e., due to the relationship with many skin diseases such as skin cancer (Chen et al. 2014); evolutionary biology i.e., due to the widely assumed environmental adaptation in skin color via positive selection (Sturm 2009); as well as anthropological and forensic applications of DNA predicting skin color of unknown individuals, including deceased modern and archaic humans, and unknown perpetrators to provide investigative leads (Kayser and de Knijff 2011). Recently, we reported a comprehensive candidate gene study identifying two genes (*BNC2* and *UGT1A*) influencing skin color variation in Europeans (Jacobs et al. 2012). To search for additional DNA variants involved in European skin color variation, the International Visible Trait Genetics (VisiGen) Consortium conducted a series of genome-wide association studies (GWASs) followed by a replication analysis in a total of 17,262 Europeans (**Table S1**) from three discovery cohorts including: the Rotterdam Study (RS) n=5,857 from the Netherlands, the Brisbane Twin Nevus Study (BTNS) n=3,459 from Australia, and the TwinsUK study, n=2,668 from the United Kingdom. Further replication was conducted in the National Child Development Study (NCDS, United Kingdom), n=5,278. Skin color phenotypes included quantitative skin color saturation (S), perceived skin darkness (PSD), the Fitzpatrick scale (FPS) of skin sensitivity to sun (Fitzpatrick 1988), and self-reported skin color darkness (note that different phenotypes were available in different cohorts, for details see **Table S1**). Functional follow-up was performed on the genomic regions identified by the skin color GWAS.

Results and Discussion

A total of five distinct genomic regions were identified that harbored DNA variants associated with skin color at genome-wide significant level ($p < 5\times10^{-8}$) including: 5p13.2 containing

SLC45A2; 6p25.3 containing *IRF4*; 15q13.1 containing *OCA2* and *HERC2*; 16q24.3 containing *MC1R*; and 20q11.22 spanning ~ 1.5Mb containing *ASIP* (**Table S2, Figure S1**). Association signals observed at all five loci were highly significantly replicated in the NCDS (p-value < 1×10^{-6}, **Table S3**).

In a multivariate analysis of the 9 top-associated single nucleotide polymorphisms (SNPs) from 9 different genes (**Table 1**), all except *SLC24A5* rs2924567 showed significant independent effects on both saturation and PSD in RS. In line with previous findings for eye (Eiberg et al. 2008; Liu et al. 2009; Liu et al. 2010; Sturm et al. 2008) and hair color (Branicki et al. 2011; Han et al. 2008). The 9 SNPs together explained a substantial proportion of the phenotypic variance depending on phenotype and population (from 3.3% saturation in RS to 16.3% for PSD in BTNS), which is much larger than typical findings from GWAS of human complex traits (e.g. hundreds of SNPs together explain 10-12% phenotypic variance of body height (Lango Allen et al. 2010)) but much smaller than that of human eye color (several SNPs explain up

Table 1. Genetic factors explaining variation of skin color measures in Europeans of the Rotterdam Study (RS, n=5,857), the Brisbane Twin Nevus Study (BTNS, n=3,459), and the TwinsUK study (n=2,668)

				Skin Saturation (RS)				PSD (RS)			PSD (BTNS)			Fitzpatric (TwinsUK)		
SNP	Gene	Chr	EA	fEA	beta	P	R2%	beta	P	R2%	fEA	beta	P	R2%	fEA	beta	P
rs183671	*SLC45A2*	5	T	0.021	0.021	4.46E-10	0.439	0.178	1.51E-09	0.482	0.029	0.505	1.64E-14	2.998	0.056	0.282	9.08E-05
rs12203592	*IRF4*	6	T	0.092	-0.011	2.26E-10	0.478	-0.055	1.64E-04	0.201	0.234	-0.220	1.09E-17	3.627	0.145	-0.157	1.12E-03
rs10756819	*BNC2*	9	G	0.332	0.005	4.40E-06	0.261	0.027	2.50E-03	0.130	0.325	0.051	3.03E-02	0.243	0.330	0.071	6.13E-03
rs1393350	*TYR*	11	A	0.235	-0.004	2.37E-04	0.163	-0.046	4.24E-06	0.304	0.292	-0.081	2.24E-03	0.573	0.282	-0.136	3.54E-07
rs17128291	*SLC24A4*	14	G	0.157	-0.003	4.24E-02	0.048	-0.038	9.55E-04	0.147	0.152	-0.067	2.95E-02	0.241	0.151	-0.044	2.05E-01
rs12913832	*HERC2*	15	A	0.200	0.008	1.08E-10	0.497	0.209	9.02E-78	5.029	0.217	0.275	1.59E-24	5.376	0.280	0.075	2.10E-02
rs2924567	*SLC24A5*	15	T	0.374	-0.002	2.11E-02	0.060	-0.016	7.34E-02	0.044	0.368	-0.017	4.78E-01	0.029	0.349	0.005	8.63E-01
rs4268748	*MC1R*	16	C	0.277	-0.009	2.59E-15	0.758	-0.049	2.15E-07	0.376	0.294	-0.158	2.84E-10	2.161	0.315	-0.214	3.62E-16
rs6059655	*RALY/ASIP*	20	A	0.079	-0.013	6.36E-13	0.608	-0.072	4.22E-06	0.278	0.099	-0.169	8.58E-06	1.057	0.096	-0.220	1.27E-07
Total R2%							3.313			6.992				16.305			

Saturation: quantitative measure derived from high-resolution digital photographs; PSD: 3-level perceived skin darkness.
For genome-wide significant hits, only the most significant SNP per gene locus (*SLC45A2*, *IRF4*, *HERC2*, *MC1R*, and *ASIP*) is included. For non-genome-wide significant hits, 4 SNPs in 4 known skin color genes are also included (*BNC2*, *TYR*, *SLC24A5*, and *SLC24A4*). EA: effect allele; fEA: frequency of the effect allele; R2%: percentage of phenotypic variance explained from multivariate analysis. In AU, the effect of sex (matched twins) and age (a small range) were excluded prior to the analysis.

to 50% phenotypic variance (Liu et al. 2010)). *HERC2* rs12913832 displayed the strongest effect, which was larger on PSD (R^2=5.02% in RS) than on saturation (R^2 = 0.497% in RS). All other highlighted SNPs explained more variance of S than PSD. SNP allele frequencies between RS and BTNS were largely similar, with the exception of *IRF4* rs12203592, for which the lighter color-associated T allele had a 2.5 fold higher frequency in BTNS (0.234) compared to that in RS (0.092). Consequently, *IRF4* rs12203592 was found as the second influential genetic factor (R^2 = 3.67%) for PSD in BTNS after *HERC2* rs12913832 (R^2 = 5.38%). These 9 SNPs were used to construct a genetically inferred skin color score in 940 samples from 54 world-wide populations (HGDP-CEPH), which showed a spatial distribution with a clear gradual increase in skin darkness from Northern Europe to Southern Europe to Northern Africa, the Middle East and Western Asia (**Figure S2**); in agreement with the known distribution of skin color across these geographic regions. Outside of these geographic regions, the inferred skin color score appeared rather similar (i.e. failing to discriminate), despite the known phenotypic skin color difference between generally lighter Asians/Native Americans and darker Africans. This demonstrates that although these 9 SNPs can explain skin color variation among Europeans, they cannot explain existing skin color differences between Asians/Native Americans and Africans. Therefore, these differences in skin color variation are likely due to different DNA variants not identifiable by this European study.

For four of the five highlighted regions i.e., 5p13.2 (**Figure 1A**), 6p25.3 (**Figure 1B**), 15q13.1 (**Figure 1C**), and 16q24.3 (**Figure 1D**), the genes responsible for the noted skin color association signals are well documented i.e., *SLC45A2*, *IRF4*, *OCA2* / *HERC2*, and *MC1R*, respectively (Liu et al. 2013). However, from previous studies it is much less clear which gene(s) amongst the 20q11.22 SNPs is/are functionally underlying the observed association (Liu et al. 2013). Unlike the other four regions, the top-associated SNP (rs6059655) in 20q11.22 (**Figure 1E**) was genome-wide significant only for quantitative skin color saturation ($p = 6.36\times10^{-13}$ in RS) and was less significant for PSD ($p = 4.22\times10^{-6}$ in RS; $p = 8.58\times10^{-6}$ in BTNS) and FPS ($p = 1.27\times10^{-7}$ in TwinsUK, **Table S2**). Among all other 20q11 SNPs with p-values smaller than 1×10^{-6}, rs1885120 within *MYH7B* and rs910873 within *PIGU* have been previously associated with melanoma risk (Brown et al. 2008). An intergenic SNP (rs4911466) has been associated with sun-burning, freckling, red hair, and skin sensitivity to sun (Sulem et al. 2008). The association signals noted at 20q11.22 span a large haplotype block of ~1.5 Mb containing 22 known genes, among which *ASIP*, a gene encoding the agouti signaling protein, and is assumed to be involved in melanogenesis (Suzuki et al. 1997). However, variants from coding regions of *ASIP* may not explain the observed association (Sulem et al. 2008). The SNP rs6059655 in intron 8 of *RALY* is ~ 182 kBp (hg19) upstream of *ASIP*. All other SNPs in this region showing association signals ($p < 1\times10^{-6}$) with skin color phenotypes were in moderate or high linkage disequilibrium with rs6059655 (LD $r^2 > 0.4$ in our European data). However, none of them displayed any significant independent association at the genome-wide level after conditioning for the rs6059655 genotype (all p-values > 0.001). Haplotype and SNP interaction analyses at 20q11 did not reveal more significant association signals for other SNPs than rs6059655 alone (**Figure S3**).

Aiming to explain the skin color association signal observed in the 20q11.22 region, we investigated the expression patterns of 22 genes in this region. First, we analyzed whole-transcriptome sequencing data obtained from 6 skin melanocytic cell lines (MCLs)

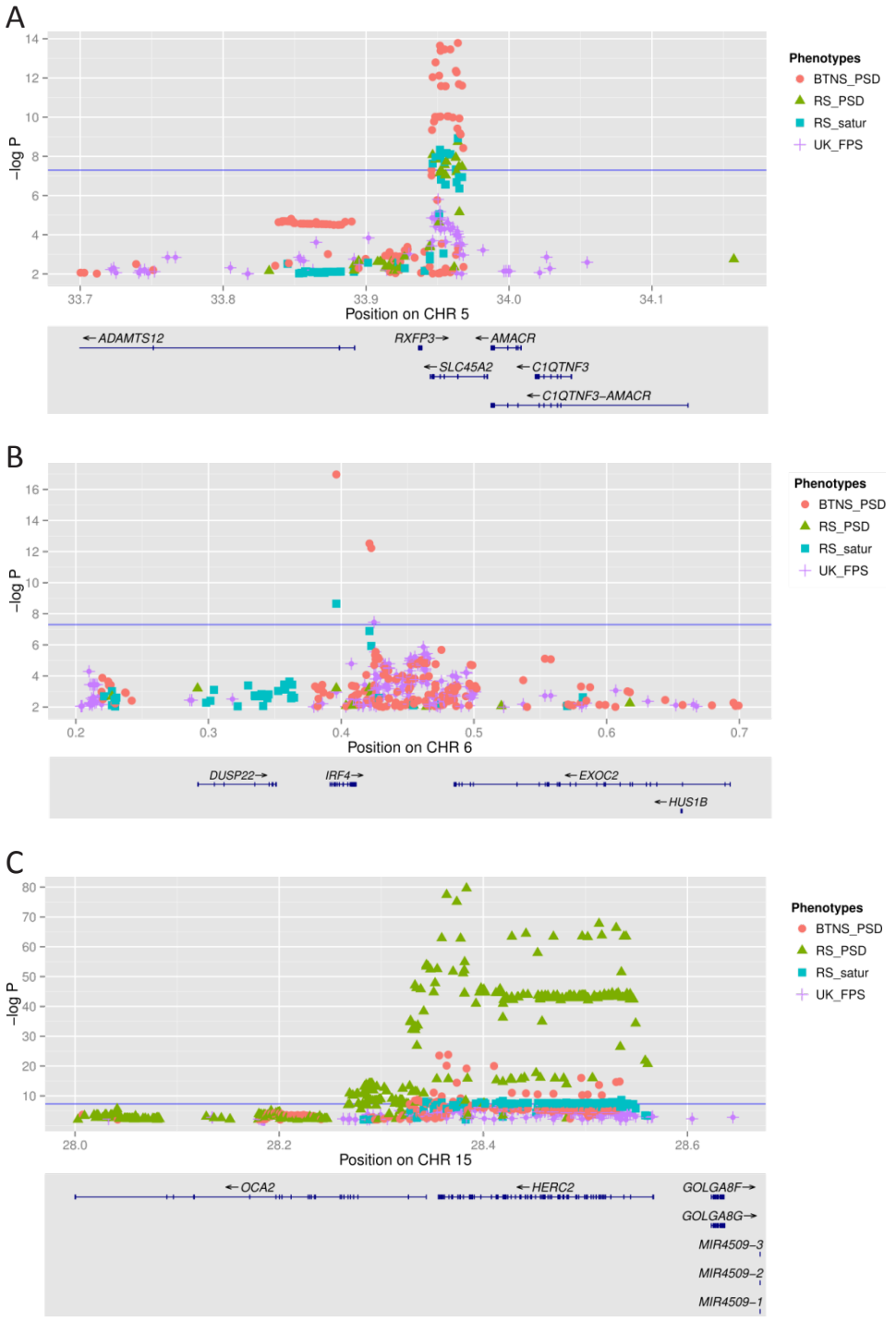


Figure 1. *(continued on next page)*

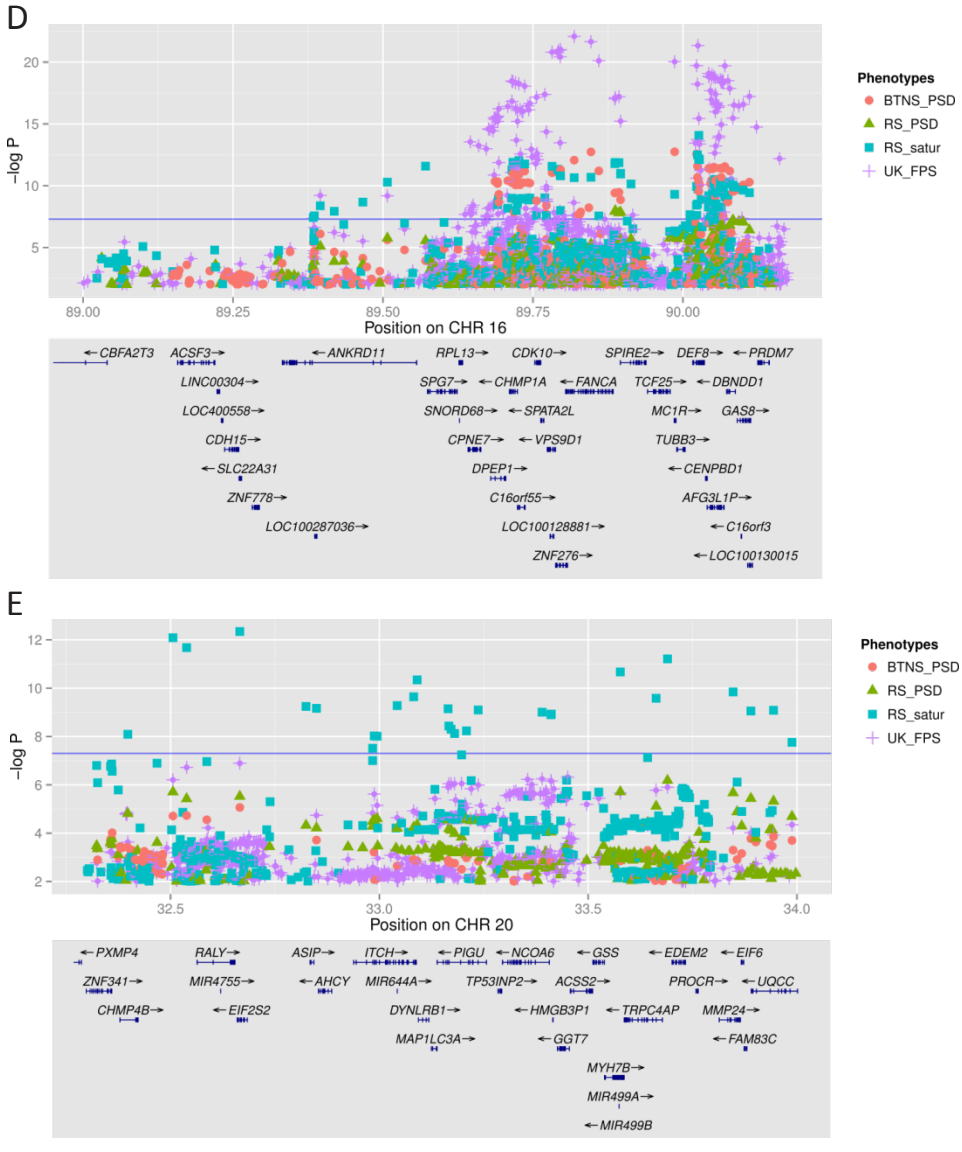


Figure 1. Regional Manhattan plots for skin color phenotypes in the Rotterdam Study, the Brisbane Twin Nevus Study, and the TwinsUK study. A. chromosome 5p13.2 (33.7-34.2 Mb) containing *SLC45A2*; B. chromosome 6p25.3 (0.2-0.7 Mb) containing *IRF4*; C. chromosome 15q13.1 (28.0-28.7 Mb) containing *OCA2* and *HERC2*; D. chromosome 16q24.3 (89.0-90.2 Mb) containing *MC1R*; and E. a large region on chromosome 20q11.22 spanning ~ 1.5Mb (32.3-34.0 Mb) containing *ASIP*. The –log10 p-values of all SNPs are plotted against their physical positions (hg19). The blue horizontal line stands for the p-value threshold of 5×10^{-8}. P-value dots are represented in colors and shapes indicating different phenotypes from different study cohorts (plink circles ➔ perceived skin darkness in BTNS, green triangles ➔ perceived skin darkness in RS, blue squares ➔ quantitative skin color saturation in RS, and purple pluses ➔ Fitzpatrick scales in TwinsUK). The physical positions of all known genes in the regions are aligned below.

Table 2. Expression profile of 22 genes at 20q11.22 in 6 melanocyte cell lines and 29 skin epidermal samples of different pigmentation status.

	melanocyte cell lines		Skin epidermal samples	rs1885120 in melanocyte cell lines	Combined
	RNA-seq	qPCR	qPCR	qPCR	p-value
RALY	+	-	++	-	0.157
EIF2S2	**++**	**+**	**++**	**-**	**0.017**
ASIP	nd	nd	nd	nd	nd
AHCY	+	-	+	-	0.312
ITCH	+	++	++	-	0.047
DYNLRB1	-	-	-	-	1.0
MAP1LC3A	+	-	-	-	0.871
PIGU	-	-	-	-	0.642
TP53INP2	-	-	-	-	0.759
NCOA6	-	++	+	-	0.083
GGT7	-	-	-	-	0.811
ACSS2	-	-	-	-	0.749
GSS	**-**	**-**	**++**	**++**	**0.0007**
MYH7B	nd	nd	nd	nd	nd
TRPC4AP	-	-	-	-	0.426
EDEM2	+	+	+	-	0.211
PROCR	-	-	+	-	0.327
MMP24	-	nd	nd	nd	nd
MMP24-AS1	-	-	-	-	1.0
EIF6	++	-	-	+	0.045
FAM83C	nd	nd	nd	nd	nd
UQCC	++	-	-	+	0.039

Transcription of 22 genes at 20q11.22 was measured in 6 melanocyte cell lines, two light pigmented (LP22, LP89), one medium pigmented (MP01), and three dark pigmented (DP74, DP80 and DP83) ones using whole-transcriptome sequencing (1st data column) and results were tested for confirmation with RT-qPCR (2nd data column). Transcription of the 22 genes in 29 skin epidermal samples from 17 light skinned and 12 dark skinned volunteers was measured using RT-qPCR (3rd data column). Correlation between rs1881520 and the expression of the 22 genes at 20q11.22 in the 6 melanocyte cell lines (4th data column). P-values of the 4 independent transcription analyses are combined to determine their significance, corrected by the number of genes tested (5th column). Correlations are denoted according to statistical significance: - ($p > 0.05$), + ($p < 0.05$), ++ ($P < 0.01$), and nd (not detected).

of two light, one moderate, and three dark pigmented individuals, followed by confirmatory analysis via reverse transcriptase quantitative PCR (RT-qPCR) of the 22 genes in the 6 MCLs as well as in 29 skin epidermal samples (SESs) derived from 12 dark and 17 light skin colored donors (detailed sample information is provided elsewhere (Visser et al. 2014)) (**Table 2**). In the whole-transcriptome sequencing data from the MCLs, we observed higher expression levels in the dark and moderate pigmented melanocyte MCLs than in the light pigmented ones for 8 of 22 genes (*RALY*, *EIF2S2*, *AHCY*, *ITCH*, *MAP1LC3A*, *EDEM2*, *EIF6* and *UQCC*). Expression of 3 genes (*ASIP*, *MYH7B* and *FAM83C*) was not detected, while for the remaining 11 genes the expression levels were not observed to be significantly different between the dark and moderate pigmented MCLs and the light pigmented MCLs. The observed differential gene expression was confirmed by RT-qPCR for 5

of the 8 genes (*RALY*, *EIF2S2*, *AHCY*, *ITCH*, and *EDEM2*) in the 6 MCLs as well as in the SESs. In addition, the genes *NCOA6* and *GSS* that were initially not highlighted by the RNA sequencing results, showed differential expression between light and dark pigmented samples in the MCLs as well as SESs for *NCOA6*, and in the SESs only for *GSS* (**Table 2,** Ct-values and p-values are indicated in **Table S4** and **Table S5**). Expression of 4 of the 22 genes (*ASIP*, *MYH7B*, *MMP24* and *FAM83C*) was not detected in the MCLs nor in the SESs (**Table 2**, Ct-values and p-values are indicated in **Table S4** and **Table S5**).

As non-coding SNPs have been reported to be involved in transcriptional regulation of nearby pigmentation genes (Guenther et al. 2014; Praetorius et al. 2013; Visser et al. 2012; Visser et al. 2014), we next investigated the correlation between the genotypes of rs6059655 and the transcription of the 22 genes at 20q11.22 in the 6 MCLs and in the 29 SESs. Both sample sets did not have enough genotypic variation for rs6059655 or its LD partners, therefore no genotype-expression correlation was observed for any of the 22 genes tested. There was, however, one exception; in the 6 MCLs, the genotypes of the LD SNP rs1885120 (LD r^2=0.82) correlated significantly with the expression of *GSS* ($p<0.01$), and potentially significantly with the expression of *EIF6* and *UQCC* ($p<0.05$) (**Table 2**, Ct-values and p-values are indicated in **Table S6**).

As we found rs1885120 and potentially rs6059655 correlating with the expression of one or more genes at 20q11.22, we further checked if the physical positions of any noted SNPs might coincide with a regional regulatory element that affects expression levels of the correlated gene(s). Using a combination of two newly obtained and several previously published ChIP-seq data sets we profiled the chromatin at 20q11.22 for features of regulatory elements (**Figure S4**). This analysis identified many different promoter and (potential) regulatory elements within 20q11.22, indicating that the region is indeed transcriptionally active in both epidermal melanocytes and epidermal keratinocytes. However, none of the regions identified with regulatory potential coincided with the physical positions of any noted SNPs.

To further investigate the expression patterns of the 22 genes at 20q11.22 and their (potential) correlation with pigmentation-SNP genotypes, we checked the expression quantitative trait locus (eQTL) data in the publically available Multiple Tissue Human Expression Resource (MuTHER) Study database (Nica et al. 2011). We found for skin biopsy samples a highly significant association between the expression of *ASIP* (HumanHT-12 array probe ILMN_1791647 targeting exon 3 of *ASIP*) and the SNPs rs1885120, rs910873 and rs17305573 (p-value = 1×10^{-26}) (**Figure 2A** and **Table S7**), while for the expression of 8 genes at 20q11.22 (*EIF2S2*, *ITCH*, *MAP1LC3A*, *GGT7*, *EDEM2*, *PROCR*, *EIF6*, and *FAM83C*of which *EIF2S2*, *ITCH*, *MAP1LC3A*, *EDEM2*, *PROCR* and *EIF6* were highlighted by at least one of our previous expression analyses more modest associations with SNP-genotypes were found, however for these 8 genes the expression-associated SNPs resulting from the eQTL analysis do not overlap with our GWAS results (**Table S7**). Notably, in the MuTHER project, full-layer skin samples were used while in our study only epidermal samples were used. The difference between our data and that obtained with the MuTHER-eQTL analysis might be explained by expression of *ASIP* in the skin dermis rather than in the melanocytes or keratinocytes located in the skin epidermis.

To test this hypothesis, we analyzed expression patterns of *ASIP* in the dermis and

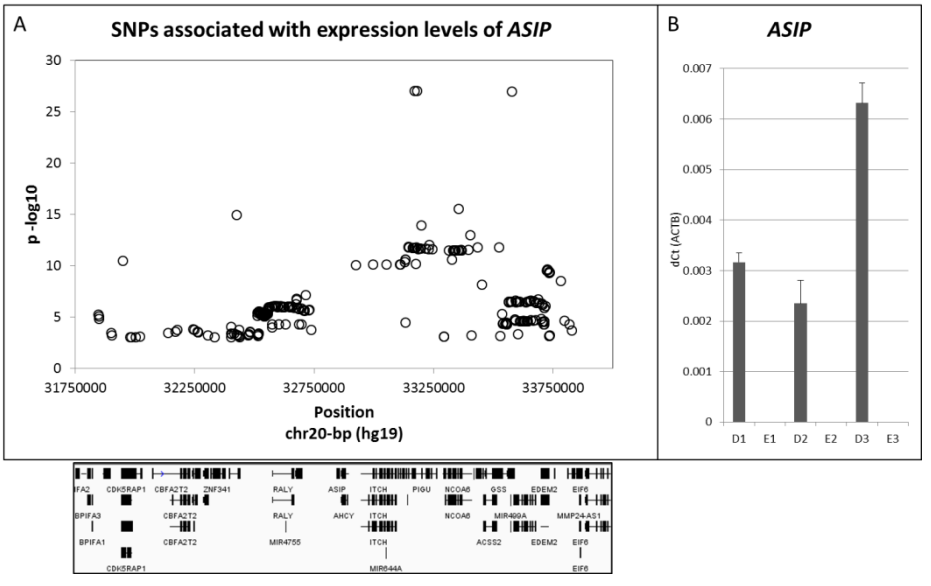


Figure 2. Expression of *ASIP* in full, dermal, and epidermal layers of skin. A. Plot of eQTL analysis on *ASIP*, where expression of *ASIP* is strongly associated with pigmentation variants rs17305573, rs910873 and rs1885120 in skin full layer biopsy samples. B. *ASIP* is exclusively expressed in the dermal layer of skin, and not in the epidermal layer of skin (nd=not detected). Samples derived from the dermal layer are denoted with 'D', samples derived from the epidermal layer are denoted with 'E'. Sample 1 has the rs1885120-CC genotype, with a dark-skin phenotype, sample 2 has the rs1885120-CC genotype, with a light-skin phenotype, sample 3 has the rs1885120-CT genotype, with a light-skin phenotype.

the epidermis of skin biopsy samples obtained from three individuals. This analysis indeed revealed a robust expression of *ASIP* in the dermal samples (not containing melanocytes), whereas in the melanocyte-containing epidermal samples *ASIP* was not detected (**Figure 2B**). In addition, expression of *ASIP* was found to be higher in the dermal sample from the rs1885120-CT heterozygote carrier than in the two samples from the rs1885120-CC homozygote carriers, which confirms the findings of the MuTHER eQTL study. It is known that not only epidermal but also dermal components contribute to normal skin pigmentation, for example dermal fibroblasts are involved in the secretion of several (paracrine) factors that modulate signaling pathways that are involved in melanocyte function and consequent skin pigmentation (Yamaguchi and Hearing 2009). As *ASIP* is a well-known antagonist of the receptor molecule MC1R located on the cell surface of melanocytes (Suzuki et al. 1997), it is possible that ASIP becomes secreted by dermal components like fibroblasts to interact with MC1R. Furthermore, it has been suggested that the epidermal melanin unit not only consists of melanocytes and keratinocytes, but also involves Langerhans cells present in the epidermis that also exist in the papillary dermis (Jimbow et al. 1991; Nordlund 2007). Several human pigmentation disorders, such as ceruloderma, a type of dermal melanosis, and forms of post-inflammatory hyperpigmentation are due to defects in the dermal layer, indicating that human pigmentation is indeed not exclusively regulated in the epidermis,

but also in the dermis (Ortonne 2012). Additionally, a melanocyte reservoir for hair and skin (re-)pigmentation consisting of melanocyte stem cells (MelSCs) is located in specific compartments of hair follicles in the dermis (Nishimura 2011). Therefore, it might also possible that *ASIP* is expressed in these MelSCs, and becomes silenced upon differentiation of the melanocytes in the epidermis.

P-values from all above experiments using were combined using Fisher's method, resulting into a list of 6 genes at 20q11.22, namely *ASIP*, *EIF2S2*, *ITCH*, *GSS*, *EIF6* and *UQCC* that are indicated with significant functional evidence to be involved in human skin color variation (combined p-value < 0.05, **Table 2**). Although the location of rs6059655, as well as previous studies on the agouti-yellow (A^y) deletion in mice (Michaud et al. 1994), suggest an involvement of the *RALY* gene in skin pigmentation, our data did not provide enough evidence supporting *RALY* as a functional human skin color gene (combined p-value = 0.15). However, its upstream-neighboring gene *EIF2S2*, of which the expression was significantly (combined p-value = 0.02) correlated with pigmentation phenotypes, was also deleted in the same A^y mutation as *RALY*, and was shown to be involved in other A^y-mutation phenotypes (Heaney et al. 2009). Moreover, the gene with the highest significance was *GSS* (combined p-value < 0.001, **Table 2**), which encodes the glutathione synthetase involved in the catalyzation of the second step of the glutathione (GSH) biosynthesis. GSH is a highly important cellular antioxidant with multiple cellular functions and major effects on melanogenesis within melanocytes. For example, GSH was shown to play a crucial role in the switching between eumelanogenesis and pheomelanogenesis by interacting with the tyrosinase enzyme (del Marmol et al. 1993) and by reacting with dopaquinone in the tyrosinase pathway (Ito 2003; Jara et al. 1988). Moreover, GSH was shown to be involved in the oxidative processes of melanin formation (Panzella et al. 2014), and was differentially detected in skin biopsy samples of different skin color (Halprin and Ohkawara 1966). Recently, an experimental study revealed that oral administration of GSH induces depigmentation of skin (Arjinpathana and Asawanonda 2012). These lines of evidence, together with our association and expression data, support an important role of *GSS* in skin coloration. Based on our analysis of the chromatin profile at the region around rs6059655, it seems unlikely that this SNP acts as an enhancer element that regulates transcription of *GSS* (or another pigmentation gene). Instead, other (yet unknown) markers in LD might do so (as was shown for *BNC2* (Visser et al. 2014)). Alternatively, rs6059655 might be involved in DNA folding or it could tag an indel.

In summary, our replicated GWAS of quantitative skin color provides the first genome-wide significant evidence for one or more common DNA variants at 20q11.22 being explicitly associated with skin coloration in Europeans. Furthermore, this study highlights additional variants associated at genome-wide significant level with skin color arising from four additional regions containing genes known to be involved in the determination of pigmentation (5p13.2 containing *SLC45A2*, 6p25.3 containing *IRF4*, 15q13.1 containing *OCA2* and *HERC2*, and16q24.3 containing *MC1R*). A combination of 9 SNPs from 9 pigmentation genes was found useful to DNA-predict skin color only in Europeans. Functional analyses prioritized two genes at 20q11.22, *EIF2S2* and *GSS*, as the most likely novel candidates responsible for the observed genetic association; both genes were significantly differentially expressed in the skin epidermis. We further showed that *ASIP* is not expressed in the epidermis containing the pigment layer, but instead in the skin dermis. Consistent with the

known biology of melanocyte cell regulation (Yamaguchi and Hearing 2009), we demonstrate that skin color is regulated not only in the pigment layer of the epidermis, but also from within the dermal layer of skin. Seen together, these findings represent a step forward in the understanding of the genetic basis of pigmentation variation in humans and are relevant to DNA phenotyping of skin color for forensic and anthropological applications.

Materials and Methods

Rotterdam Study (RS)

The Dutch European RS (Hofman et al. 2013) is a population-based prospective study consisting of a main cohort and two extensions. The RS is ongoing since 1990 and currently includes 14,926 participants living in a particular suburb of Rotterdam in the Netherlands. The Medical Ethics Committee of the Erasmus University Medical Center approved the study protocol and all participants provided written informed consent. The current study includes 5,857 participants of Northwestern European ancestry, with microarray genotype data and digital photographs available. No exclusions have been made on skin related diseases. We used the same set of photographs as obtained and described in our previous eye color GWAS (Liu et al. 2010). Skin color phenotypes were derived and described in detail in a previous study (Jacobs et al. 2012); here we focus on skin saturation (mean 0.524, sd 0.063, min 0.278, max 0.808) derived from digital photos and 3-level perceived skin darkness graded by a dermatologist (very white 14.84%, white 73.68%, white-to-olive 11.48%). Microarray genotyping was conducted using the Infinium II HumanHap550K and Human 610 Quad Arrays of Illumina. Details on genotyping and quality controls are described elsewhere (Liu et al. 2010). Genotypes were imputed using the 1000-Genomes Project as the reference panel (Phase 1, integrated variant set across 1,092 individuals, v2, March 2012) using the MaCH and minimac software packages. After all quality controls (MAF > 0.01, marker call rate > 0.97, and HWE > 1×10-6), the final data set included 11,155,022 SNPs with imputation Rsq > 0.4.

Brisbane Twin Nevus Study (BTNS)

The Australian BTNS has recruited adolescent twins, their siblings and parents over the past 22 years into an ongoing study of genetic and environmental factors contributing to the development of pigmented nevi and other risk factors for skin cancer. The proband twins are recruited at age twelve years via schools around Brisbane, Australia, and followed up at age fourteen. The sample is of Northern European origin (mainly Anglo-Celtic >95%). All cases and controls gave informed consent to participation in this study, and the study protocol was approved by appropriate institutional review boards. Skin color at age 12–14 years was reported by participants as one of three categories: fair/light, medium, or olive/dark.

DNA samples from the BTNS were genotyped by the Scientific Services Division at deCODE Genetics, Iceland using the Illumina 610-Quad BeadChip; genotypes were called with the Illumina BeadStudio software. For the GWAS, we first applied filters to SNP data before evaluating genotyping quality per individual and excluded SNPs with a mean BeadStu-

dio GenCall score < 0.7. Next, we excluded poorly performing samples (call rate < 0.95) and SNPs with call rate < 0.95, Hardy-Weinberg equilibrium P < 10-6, or minor allele frequency <0.01. Following these exclusions, we compared self -reported with genotype inferred family relationships, the latter based on genome-wide IBS sharing. Forty eight families with pedigree errors were identified; 21 samples from these families were excluded to correct errors which could not be resolved. No SNPs or individuals showed segregation patterns inconsistent with Mendelian inheritance in >5% of families and SNPs, respectively. Lastly, we excluded 88 individuals identified as outliers from populations of European descent through the estimation of genetic ancestry using EIGENSTRAT and data from eleven populations of the HapMap 3 and five Northern European populations genotyped by the GenomeEUtwin consortium. Following these exclusions, there remained 529,721 SNPs and 4,296 individuals with genotype data for analysis. Imputation was undertaken with the use of the phased data from the HapMap samples of European ancestry (CEU; build 36, release 22) and MACH. After imputation quality controls (MACH Rsq>0.4), this dataset included 2,558,980 SNPs.

TwinsUK

The TwinsUK study included 2,668 phenotyped participants (97% female and all of Caucasian ancestry) within the TwinsUK adult twin registry based at St. Thomas' Hospital in London. Twins largely volunteered unaware of the skin research interests at the time of enrolment and gave fully informed consent under a protocol reviewed by the St. Thomas' Hospital Local Research Ethics Committee. Genotyping of the TwinsUK cohort was done with a combination of Illumina HumanHap300 and HumanHap610Q chips. Intensity data for each of the arrays were pooled separately and genotypes were called with the Illuminus32 calling algorithm, thresholding on a maximum posterior probability of 0.95 as previously described (Small et al. 2011). Imputation was performed using the IMPUTE 2.0 software package (https://mathgen.stats.ox.ac.uk/impute/impute.html) using haplotype information from the 1000 Genomes Project (Phase 1, integrated variant set across 1,092 individuals, v2, March 2012). Imputed genotypes were subsequently converted into a MACH format (http://www.sph.umich.edu/csg/abecasis/MACH/tour/input_files.html) and analyzed with mach2qtl (http://www.sph.umich.edu/csg/abecasis/MACH/download/mach2qtl.source.V112.tgz).

National Child Development Study (NCDS)

The NCDS is a cohort study of 17000 people born in England, Scotland and Wales in a single week of 1958. The participants have been extensively phenotyped on multiple occasions, including a biomedical survey, which was designed to obtain objective measures of ill-health and biomedical risk factors in order to address a wide range of specific hypotheses relating to anthropometry: cardiovascular, respiratory and allergic diseases; visual and hearing impairment; and mental ill-health. In 2003, as part of the biomedical survey, 9377 participants completed an item on skin colour, reporting it on a scale of "light", "medium" or "dark". Individuals were genotyped on both the Immunochip and Metabochip disease centred SNP arrays. SNPs set were combined, data from duplicated were SNP merged, and monomorphic

SNPs, SNPs exhibiting Hardy-Weinberg disequilibrium (P < 1e-6) or SNPs with genotyping failure rate < 0.98 were removed. A total of 298548 SNPs were then available. Imputation for the regions of interest was performed using IMPUTE2 and the 1000 Genomes Phase1 phased dataset v3 dated 2010-11-23, and the reference set haplotypes estimated using SHAPEIT2 (ALL.integrated_phase1_SHAPEIT_16-06-14). The current study included 5,278 NCDS participants for whom both skin colour and genotype data were available.

Statistical genetic analyses

GWAS were conducted in RS using linear regression assuming additive genetic effect and adjusted for sex, age, and 4 main dimensions from MDS analysis, where p-values equal or smaller than 5×10-8 were considered to be genome-wide significant. A next round GWAS was conducted conditioning on the genotypes of significant SNPs from a previous GWAS until no more significant SNPs could be identified. Inflation factors were estimated as 1.015 for skin saturation and 1.011 for perceived skin darkness and were adjusted using the genomic control method. The GWAS in TwinsUK cohort was conducted using mach2qtl v1.12 (http://www.sph.umich.edu/csg/abecasis/MACH/download/mach2qtl.source.V112.tgz). The genomic inflation factor was 1.01 for the Fitzpatrick scale GWAS.

Genome-wide Manhattan and Q-Q plots were generated using R scripts from. Regional Manhattan plots were constructed using software package locus zoom (Pruim et al. 2010). To access the overall genetic contribution on skin coloration, we conducted a multivariate analysis including 9 DNA variants from 9 genes, i.e., 5 highlighted in the present study including RALY rs6059655, HERC2 rs12913832, IRF4 rs12203592, SLC45A2 rs183671, MC1R rs4268748 and 4 suggested in previous studies (Han et al. 2008; Jacobs et al. 2012; Lamason et al. 2005; Sulem et al. 2008) including BNC2 rs10756819, TYR rs1393350, SLC24A4 rs17128291, and SLC24A5 rs2924567 (Table 1). Since both quantitative skin color saturation and the 3-level PSD phenotypes were available in RS, the genetic effects on these two phenotypes could be compared. The multivariate analysis including sex, age, and 9 SNPs from 9 genes were conducted in RS and BTNS in an iterative manner to access the R-squared change due to individual factors using R scripting. Base on this multivariate analysis we further inferred a skin color score for 940 samples from 54 populations in the HGDP database (http://www.cephb.fr/en/hgdp/diversity.php) using the sum of the number of darker skin-associated alleles weighted by the regression betas for skin saturation. Conditional analyses were conducted for all associated regions conditioning on the genotype status of the top-associated SNP. Haplotype analyses were conducted using R library haplo.stats. Collapsed double heterozygosity analyses in all of the associated regions were conducted as previously described (Liu et al. 2011). SNP interaction analyses were conducted between SNPs in the MC1R region and the ASIP region using a previous described F statistic (Liu et al. 2010). Gene transcriptions were compared between genotype carriers (wild-type vs. others) using a t-test. A combined p-value was derived for each gene by combining p-values from k independent experiments using Fisher's combined probability test, i.e., , which is relatively conservative due to accumulation of df's.

Functional genetic analyses

We investigated expression patterns of 22 genes located within the RALY-UQCC region in 6 human skin melanocytic cell lines derived from donors with different skin color (lightly pigmented LP22 and LP89, moderately pigmented MP01 and darkly pigmented DP74, DP80 and DP83), in a set of 29 skin samples derived from donors with either light (n=17) or dark (n=12) skin pigmentation. Left-over patient skin material was collected under informed consent and with approval from the Medical Ethics Committee (METC) of Erasmus MC. Details about the cell lines, the skin samples and the methods have been described previously(Visser et al. 2014). In brief; the cell lines were grown following the manufacturer's instructions (Cascade Biologics, Invitrogen), RNA and DNA were co-extracted using TriPure Isolation Reagent, followed by a purification step (OneStep™ PCR Inhibitor Removal Kit, Zymo Research Corporation) to remove melanin. The skin epidermal and dermal samples were obtained by separating the epidermal layer from surgically-removed skin biopsies, RNA and DNA were co-extracted using Qiagen Allprep mini kit, followed by the above described purification step to remove melanin. The reverse-transcriptase (RT) reaction was performed using RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas GmbH) according to the manufacturer's instructions. Quantitative real-time PCR reactions for gene-expression analysis were performed using the iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories). RNA sequencing was performed using a PGM (Life Technologies). RNA samples obtained from the 6 melanocyte cell lines were first treated with RiboMinus Eukaryote kit v2 (Life Technologies) to remove rRNA, after which the whole-transcriptome libraries were constructed using the Ion Total-RNA Seq Kit v2 (Life Technologies). Snapshot analysis was used to genotype the skin-color associated SNPs. Primer sequences are available on request.

We profiled the chromatin of region 20q11.22 spanning the 22 genes (RALY-UQCC) harboring the identified associated skin-color SNPs. We considered several data sets that represent features associated with regulatory regions: ChIP-seq analysis in a lightly pigmented melanocytic cell line (LP22), a darkly pigmented melanocytic cell line (DP74) (Palstra et al, manuscript in preparation), and in a normal human epidermal keratinocytic cell line (NHEK(Rosenbloom et al. 2013)) of acetylated histone H3 (H3K27Ac), an active chromatin mark (Creyghton et al. 2010), DNaseI hypersensitive sites in epidermal skin melanocytes and in the NHEK cell line (Rosenbloom et al. 2013); ChIP-seq data for the transcription factor MITF in melanocytic cells (Strub et al. 2011), MITF is the melanocyte master regulator(Levy et al. 2006), ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1(-Li et al. 2012), a ubiquitously expressed transcription factor that was reported to play an important role in melanocyte development by interacting with the melanocyte-specific isoform of MITF(Li et al. 2012); predicted melanocyte-specific enhancers (Gorkin et al. 2012) and Phastcons conserved elements inferred from 46 way alignments of placental mammals (Siepel et al. 2005).

Acknowledgements

We thank all participants of the Rotterdam Study, the Brisbane Twin Nevus Study, Twins UK Study, and the National Child Development Study.

Author contributions

Conceived and designed the experiments: FL MV RJTSP MK. Analyzed the data: FL MV DLD OL KZ. Data preparation: PGH SW LC AW GZ GWM AKH MM DG VB FR AH WFJvI AGU TDS NGM TECN. Wrote the manuscript: FL MV MK with contributions from RAS FR AGU RJTSP TDS NGM TECN. All authors critically reviewed and approved the manuscript.

Funding

This work was supported in part by the Erasmus MC University Medical Center Rotterdam and funds from the Netherlands Genomics Initiative/Netherlands Organization of Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands. The Rotterdam Study is funded by Erasmus MC University Medical Center Rotterdam and Erasmus University Rotterdam; the Netherlands Organization for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE2); the Ministry of Education, Culture, and Science of the Netherlands; the Ministry for Health, Welfare, and Sports of the Netherlands; the European Commission (DG XII); and the Municipality of Rotterdam. The generation and management of GWAS genotype data for the Rotterdam Study was supported by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012); the Research Institute for Diseases in the Elderly (014-93-015; RIDE2); and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) project 050-060-810. The Brisbane Twin Nevus Study (BTNS) is funded by the National Institutes of Health, United States (HD50735), and the National Health and Medical Research Council (NHMRC), Australia (496682). The Twins UK study is funded by the Wellcome Trust, European Community's Seventh Framework Program (FP7/2007-2013)/grant agreement HEALTH-F2-2008-201865-GEFOS and (FP7/2007-2013), ENGAGE project grant agreement HEALTH-F4-2007-201413, and the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The Twins UK study also receives support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR Senior Investigator. The Twins UK study also received support from a Biotechnology and Biological Sciences Research Council (BBSRC) project grant (G20234) and a U.S. National Institutes of Health (NIH)/National Eye Institute (NEI) grant (1RO1EY018246), and genotyping was supported by the NIH Center for Inherited Disease Research. The Twins UK study also received support from the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's and St. Thomas' National Health Service Foundation Trust partnering with King's College London. The NCDS biomedical survey was funded under the MRC 'Health of the Public' initiative, and was carried out in collaboration with the Institute of Child Health, St George's Hospital Medical School, and NatCen. Identitas Inc. previously sponsored the VisiGen Consortium via a research grant to a number of the respective academic institutions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interest

TS an MK have consulted for Identitas Inc. and are on the SAB but without financial or other benefits. All other authors declared that no competing interests exist.

References

Arjinpathana N, Asawanonda P (2012) Glutathione as an oral whitening agent: a randomized, double-blind, placebo-controlled study. J Dermatolog Treat 23: 97-102. doi: 10.3109/09546631003801619

Branicki W, Liu F, van Duijn K, Draus-Barini J, Pospiech E, Walsh S, Kupiec T, Wojas-Pelc A, Kayser M (2011) Model-based prediction of human hair color using DNA variants. Hum Genet 129: 443-54. doi: 10.1007/s00439-010-0939-8

Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, Thomas S, Schmid H, Holland EA, Gillanders EM, Duffy DL, Maskiell JA, Jetann J, Ferguson M, Stephan DA, Cust AE, Whiteman D, Green A, Olsson H, Puig S, Ghiorzo P, Hansson J, Demenais F, Goldstein AM, Gruis NA, Elder DE, Bishop JN, Kefford RF, Giles GG, Armstrong BK, Aitken JF, Hopper JL, Martin NG, Trent JM, Mann GJ, Hayward NK (2008) Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet 40: 838-40. doi: 10.1038/ng.163

Chen H, Weng QY, Fisher DE (2014) UV signaling pathways within the skin. Journal of Investigative Dermatology 134: 2080-5. doi: 10.1038/jid.2014.161

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107: 21931-6. doi: 10.1073/pnas.1016071107

del Marmol V, Solano F, Sels A, Huez G, Libert A, Lejeune F, Ghanem G (1993) Glutathione depletion increases tyrosinase activity in human melanoma cells. Journal of Investigative Dermatology 101: 871-4.

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 Genet 123: 177-87. doi: 10.1007/s00439-007-0460-x

Fitzpatrick TB (1988) The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol 124: 869-71.

Gorkin DU, Lee D, Reed X, Fletez-Brant C, Bessling SL, Loftus SK, Beer MA, Pavan WJ, McCallion AS (2012) Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. Genome Res 22: 2290-301. doi: 10.1101/gr.139360.112

Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM (2014) A molecular basis for classic blond hair color in Europeans. Nat Genet. doi: 10.1038/ng.2991

Halprin KM, Ohkawara A (1966) Glutathione and human pigmentation. Arch Dermatol 94: 355-7.

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4: e1000074. doi: 10.1371/journal.pgen.1000074

Heaney JD, Michelson MV, Youngren KK, Lam MY, Nadeau JH (2009) Deletion of eIF2beta suppresses testicular cancer incidence and causes recessive lethality in agouti-yellow mice. Hum Mol Genet 18: 1395-404. doi: 10.1093/hmg/ddp045

Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, Klaver CC, Nijsten TE, Peeters RP, Stricker BH, Tiemeier HW, Uitterlinden AG, Vernooij MW (2013) The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol 28: 889-926. doi: 10.1007/s10654-013-9866-z

Ito S (2003) The IFPCS presidential lecture: a chemist's view of melanogenesis. Pigment Cell Res 16: 230-6.

Jacobs LC, Wollstein A, Lao O, Hofman A, Klaver CC, Uitterlinden AG, Nijsten T, Kayser M, Liu F (2012) Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. Hum Genet 132: 147-158. doi: 10.1007/s00439-012-1232-9

Jara JR, Aroca P, Solano F, Martinez JH, Lozano JA (1988) The role of sulfhydryl compounds in mammalian melanogenesis: the effect of cysteine and glutathione upon tyrosinase and the intermediates of the pathway. Biochim Biophys Acta 967: 296-303.

Jimbow K, Salopek TG, Dixon WT, Searles GE, Yamada K (1991) The epidermal melanin unit in the

pathophysiology of malignant melanoma. Am J Dermatopathol 13: 179-88.
Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12: 179-92. doi: 10.1038/nrg2952
Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310: 1782-6. doi: 10.1126/science.1116238
Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Carola Zillikens M, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Hua Zhao J, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, Konig IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832-8. doi: 10.1038/nature09410
Levy C, Khaled M, Fisher DE (2006) MITF: master regulator of melanocyte development and melanoma oncogene. Trends Mol Med 12: 406-14. doi: 10.1016/j.molmed.2006.07.008
Li J, Song JS, Bell RJ, Tran TN, Haq R, Liu H, Love KT, Langer R, Anderson DG, Larue L, Fisher DE (2012) YY1 regulates melanocyte development and function by cooperating with MITF. PLoS Genet 8: e1002688. doi: 10.1371/journal.pgen.1002688
Liu F, Struchalin MV, van Duijn K, Hofman A, Uitterlinden AG, van Duijn C, Aulchenko YS, Kayser M (2011) Detecting Low Frequent Loss-of-Function Alleles in Genome Wide Association Studies with Red Hair Color as Example. PLoS One 6: e28145. doi: ARTN e28145
DOI 10.1371/journal.pone.0028145
Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens ACJW, Kayser M (2009) Eye color and the prediction of complex phenotypes from genotypes. Current Biology 19: R192-R193. doi: DOI 10.1016/j.cub.2009.01.027
Liu F, Wen B, Kayser M (2013) Colorful DNA polymorphisms in humans. Semin Cell Dev Biol 24: 562-575. doi: DOI 10.1016/j.semcdb.2013.03.013
Liu F, Wollstein A, Hysi PG, Ankra-Badu GA, Spector TD, Park D, Zhu G, Larsson M, Duffy DL, Montgomery GW, Mackey DA, Walsh S, Lao O, Hofman A, Rivadeneira F, Vingerling JR, Uitterlinden AG, Martin NG, Hammond CJ, Kayser M (2010) Digital Quantification of Human Eye Color Highlights Genetic Association of Three New Loci. PLoS Genet 6: e1000934. doi: ARTN e1000934
DOI 10.1371/journal.pgen.1000934
Michaud EJ, Bultman SJ, Klebig ML, van Vugt MJ, Stubbs LJ, Russell LB, Woychik RP (1994) A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (Ay) mutation. Proc Natl Acad Sci U S A 91: 2562-6.
Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, Hedman AK, Bataille V, Tzenova Bell J, Surdulescu G, Dimas AS, Ingle C, Nestle FO, di Meglio P, Min JL, Wilk A, Hammond CJ, Hassanali N, Yang TP, Montgomery SB, O'Rahilly S, Lindgren CM, Zondervan KT, Soranzo N, Barroso I, Durbin R, Ahmadi K, Deloukas P, McCarthy MI, Dermitzakis ET, Spector TD (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 7: e1002003. doi: 10.1371/journal.pgen.1002003
Nishimura EK (2011) Melanocyte stem cells: a melanocyte reservoir in hair follicles for hair and skin pigmentation. Pigment Cell Melanoma Res 24: 401-10. doi: 10.1111/j.1755-148X.2011.00855.x
Nordlund JJ (2007) The melanocyte and the epidermal melanin unit: an expanded concept. Dermatol Clin 25: 271-81, vii. doi: 10.1016/j.det.2007.04.001
Ortonne JP (2012) Normal and abnormal skin color. Ann Dermatol Venereol 139 Suppl 4: S125-9. doi: 10.1016/S0151-9638(12)70123-0
Panzella L, Leone L, Greco G, Vitiello G, D'Errico G, Napolitano A, d'Ischia M (2014) Red human hair

pheomelanin is a potent pro-oxidant mediating UV-independent contributory mechanisms of melanomagenesis. Pigment Cell Melanoma Res 27: 244-52. doi: 10.1111/pcmr.12199

Praetorius C, Grill C, Stacey SN, Metcalf AM, Gorkin DU, Robinson KC, Van Otterloo E, Kim RS, Bergsteinsdottir K, Ogmundsdottir MH, Magnusdottir E, Mishra PJ, Davis SR, Guo T, Zaidi MR, Helgason AS, Sigurdsson MI, Meltzer PS, Merlino G, Petit V, Larue L, Loftus SK, Adams DR, Sobhiafshar U, Emre NC, Pavan WJ, Cornell R, Smith AG, McCallion AS, Fisher DE, Stefansson K, Sturm RA, Steingrimsson E (2013) A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. Cell 155: 1022-33. doi: 10.1016/j.cell.2013.10.022

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26: 2336-7. doi: 10.1093/bioinformatics/btq419

Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ (2013) ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res 41: D56-63. doi: 10.1093/nar/gks1172

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15: 1034-50. doi: 10.1101/gr.3715005

Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A, Thorsteindottir U, Shin SY, Richards HB, Soranzo N, Ahmadi KR, Lindgren CM, Stefansson K, Dermitzakis ET, Deloukas P, Spector TD, McCarthy MI (2011) Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet 43: 561-4. doi: 10.1038/ng.833

Strub T, Giuliano S, Ye T, Bonet C, Keime C, Kobi D, Le Gras S, Cormont M, Ballotti R, Bertolotto C, Davidson I (2011) Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. Oncogene 30: 2319-32. doi: 10.1038/onc.2010.612

Sturm RA (2009) Molecular genetics of human pigmentation diversity. Hum Mol Genet 18: R9-17. doi: 10.1093/hmg/ddp003

Sturm RA, Duffy DL, Zhao ZZ, Leite FP, 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 Genet 82: 424-31. doi: 10.1016/j.ajhg.2007.11.005

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G, Palsson S, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Vermeulen SH, Goldstein AM, Tucker MA, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K (2008) Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40: 835-7. doi: 10.1038/ng.160

Suzuki I, Tada A, Ollmann MM, Barsh GS, Im S, Lamoreux ML, Hearing VJ, Nordlund JJ, Abdel-Malek ZA (1997) Agouti signaling protein inhibits melanogenesis and the response of human melanocytes to alpha-melanotropin. Journal of Investigative Dermatology 108: 838-42.

Visser M, Kayser M, Palstra RJ (2012) HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res 22: 446-55. doi: 10.1101/gr.128652.111

Visser M, Palstra RJ, Kayser M (2014) Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. Hum Mol Genet. doi: 10.1093/hmg/ddu289

Yamaguchi Y, Hearing VJ (2009) Physiological factors that regulate skin pigmentation. Biofactors 35: 193-9. doi: 10.1002/biof.29

Supplemental figures and tables

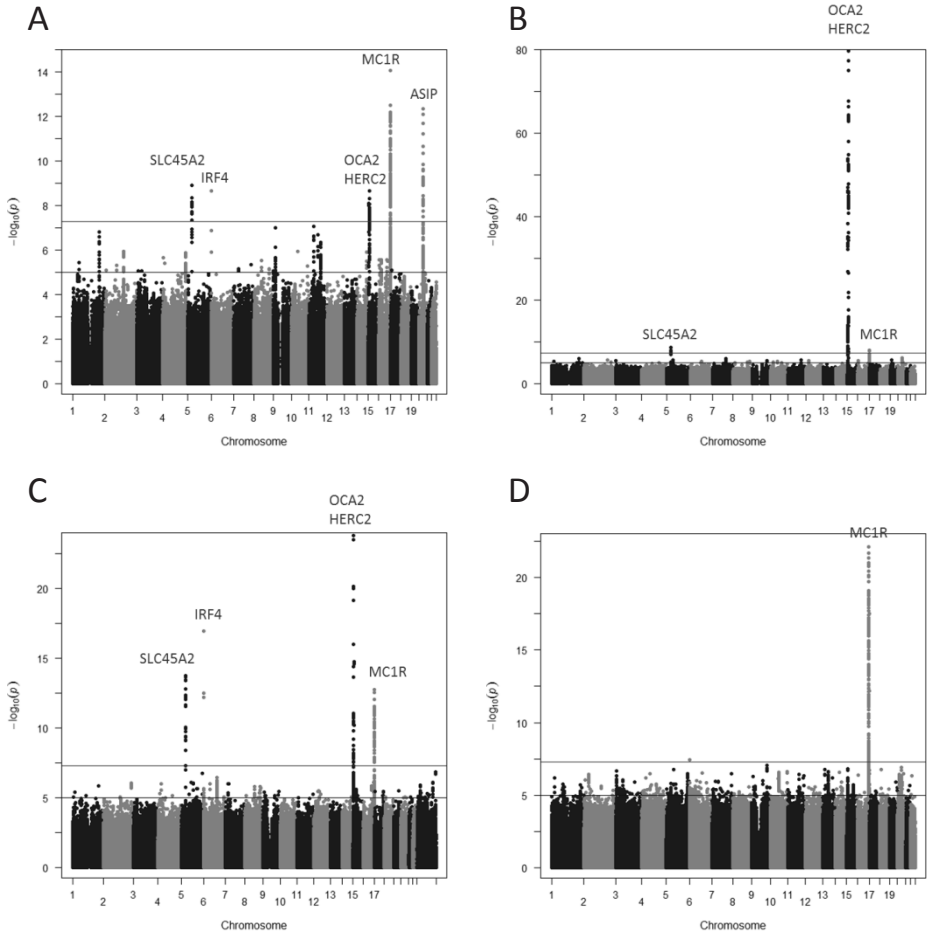


Figure S1. Manhattan plot of GWAS results for skin color phenotypes in Rotterdam Study, Brisbane Twin Nevus Study, and TwinsUK study. A, quantitative skin color saturation extracted from digital photos in the Rotterdam Study (RS, n=5,857); B, 3-level (very white, white, white-to-olive) perceived skin darkness in Rotterdam Study (RS, n=5,857); C, 3-level (fair/light, medium, or olive/dark) perceived skin darkness in the Brisbane Twin Nevus Study (BTNS, n=3,456); D, Fitzpatrick scale of sensitivity to sun (6 levels) in the TwinsUK study (n=2,668). The –log10 p-values of all SNPs are plotted against their physical positions over the genome (hg19). The blue and red horizontal lines stand for the p-value thresholds of 1×10^{-5} and 5×10^{-8}, respectively. Known pigmentation genes in the regions showing significant (p-value$<5\times10^{-8}$) association are highlighted in red color.

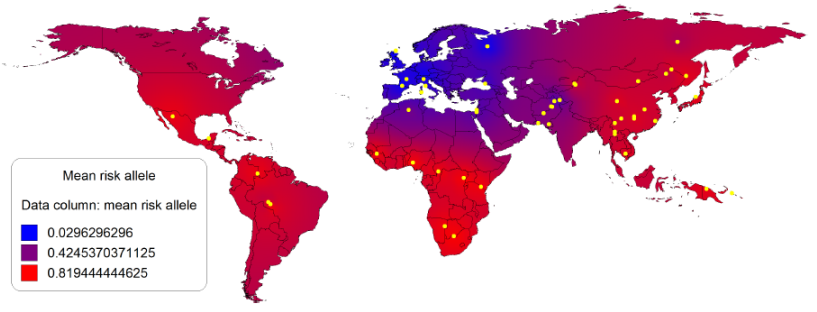


Figure S2. Spatial distribution of a genetically inferred skin color score in 940 samples from 54 populations of the HGDP-CEPH. The skin color score for 940 world-wide subjects was calculated as the sum of the number of darker skin-associated alleles weighted by the regression betas for saturation using 9 SNPs from 9 gene regions (see table 1). Yellow dots represent the geographic location of the HGDP-CEPH population samples.

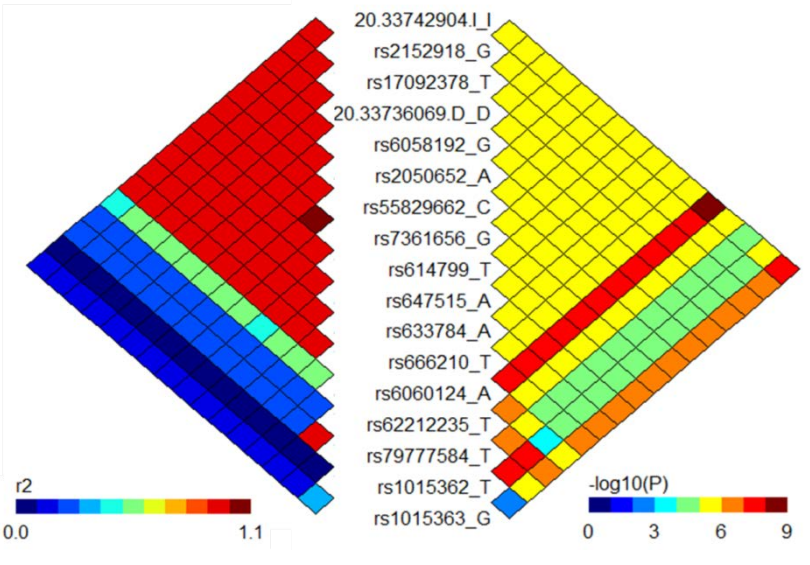


Figure S3. Haplotypes associated with skin color saturation between 3,184 SNPs on 20q11.2. A total of 3,184 SNPs within a large region (32.3-34.0 Mb) on chromosome 20q11.2 were tested in a pair-wise manner for haplotype association with skin color saturation in the Rotterdam Study. A total of 17 SNPs with at least one P values $<1\times10^{-7}$ are shown. Left part: LD r2, right part: significance of pair-wise haplotype association.

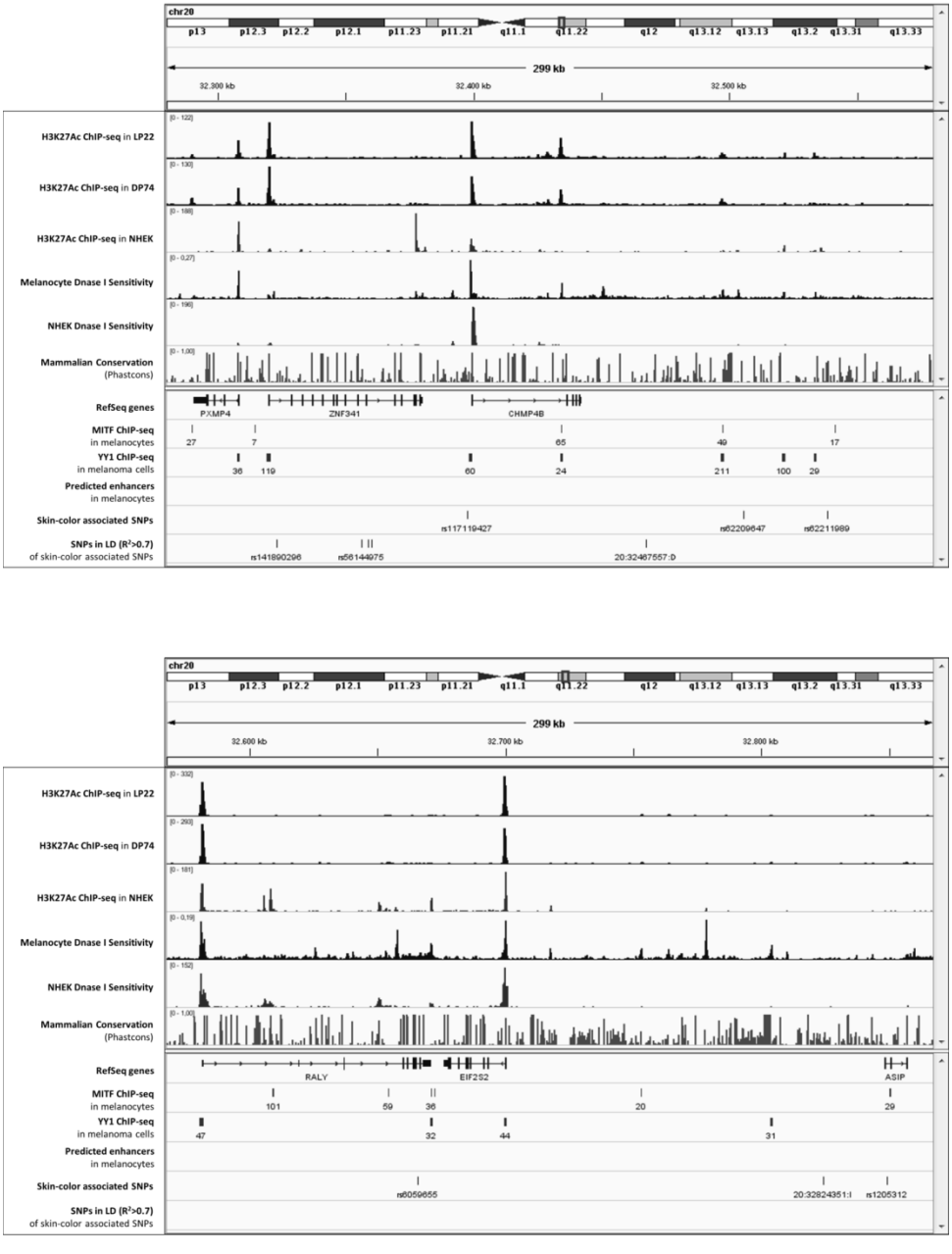


Figure S4. *(continued on next page)*

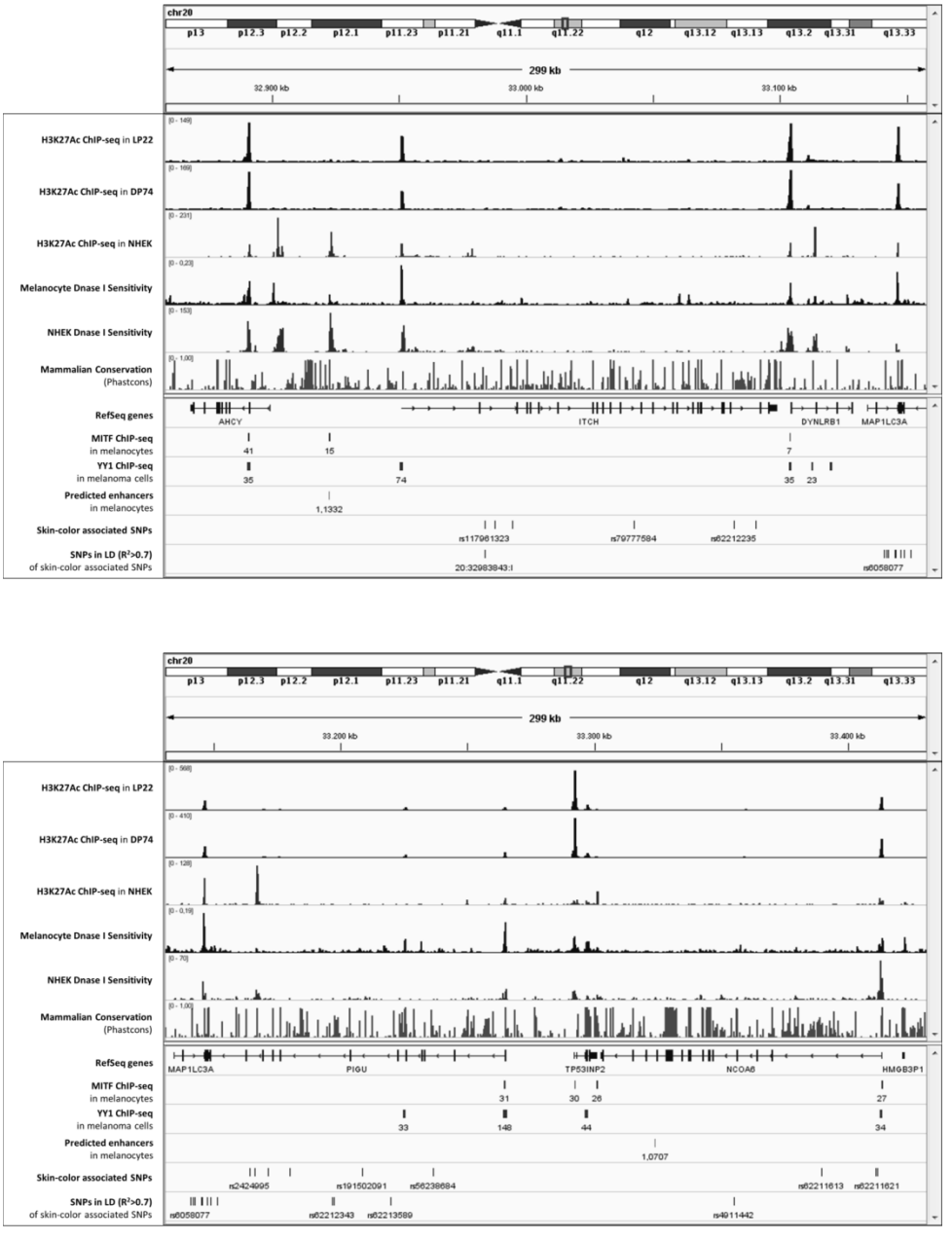


Figure S4. *(continued on next page)*

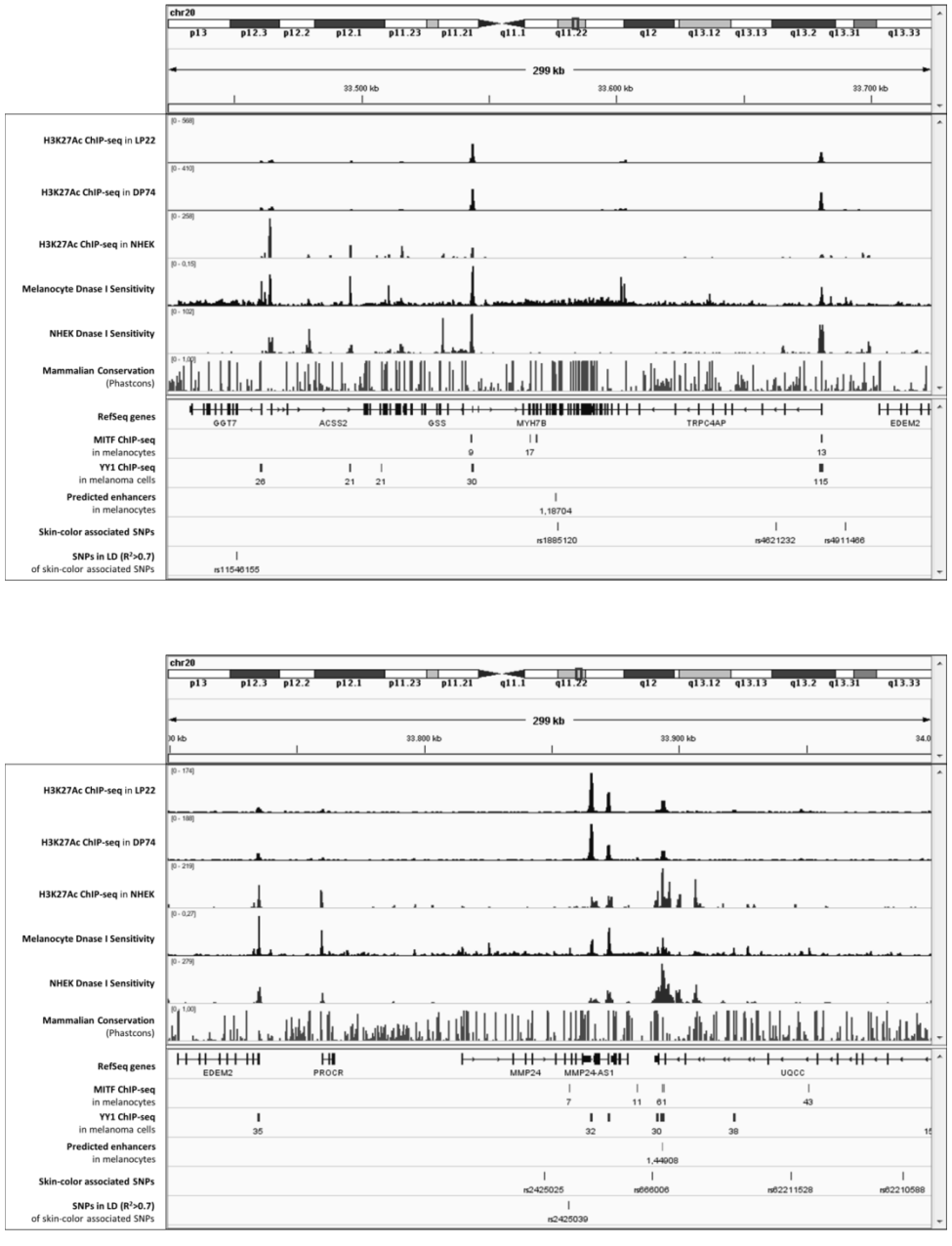


Figure S4. *(continued on next page)*

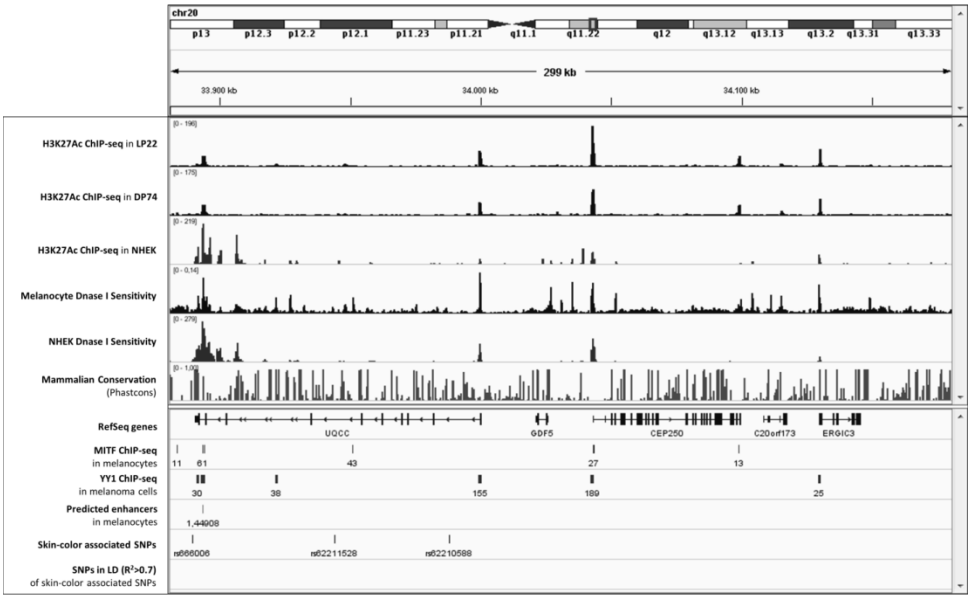


Figure S4. Chromatin profile of 22 genes on 20q11.22. IGV genome browser shows a 1.9 mb window at 20q11.22 including 22 genes (from *RALY* to *UQCC*) containing the skin-color association signals in this region. To investigate the chromatin for features of enhancer elements, the following tracks are included: ChIP-seq analysis in a lightly pigmented melanocytic cell line (LP22), a darkly pigmented melanocytic cell line (DP74) (Palstra et al, manuscript in preparation), and in a normal human epidermal keratinocytic cell line of acetylated histone H3 (H3K27Ac), an active chromatin mark, DNaseI hypersensitive sites in epidermal skin melanocytes and in the NHEK cell line; ChIP-seq data for the transcription factor MITF in melanocytic cells, MITF is the melanocyte master regulator, ChIP-seq data in MALME-3M melanoma cells for the transcription factor YY1, a ubiquitously expressed transcription factor that was reported to play an important role in melanocyte development by interacting with the melanocyte-specific isoform of MITF; predicted melanocyte-specific enhancers and Phastcons conserved elements inferred from 46 way alignments of placental mammals. See the method section for details about data sources.

Table S1. Study cohorts

Rotterdam Study (N=5857)				
	Mean	*SD*	*Min*	*Max*
Skin color saturation	0,524	0,063	0,278	0,808
Perceived skin darkness	*N*	*%*		
Light	869	14,84		
Fair	4315	73,67		
Fair-to-olive	673	11,49		
Australian (N=3459)				
Perceived skin darkness	*N*	*%*		
Light	1480	42,79		
Fair	1654	47,82		
Olive	325	9,40		
TwinsUK (N=2868)				
Fitzpatrick Scales	*N*	*%*		
1	364	12,69		
2	932	32,5		
3	1041	36,3		
4	421	14,68		
5	49	1,71		
6	61	2,13		
National Child Development Study (N=5278)				
Self-report skin color	*N*	*%*		
Light	3608	68,36		
Medium	1656	31,38		
Dark	14	0,27		

Table S2. *(available on request)*

Table S3. Replication results in National Study (N=5268)

SNP	Chr	Pos(Mbp)	Alleles	A1	Freq(A1)	P-value
rs16891982	5	33,99	C/G	G	0,0367	3,00E-17
rs35412	5	34,00	C/G	C	0,0185	2,00E-14
rs12203592	6	0,34	A/G	A	0,2236	4,00E-15
rs1393350	11	88,65	A/G	A	0,2918	9,00E-08
rs12913832	15	26,04	A/G	A	0,221	2,00E-15
rs916977	15	26,19	A/G	A	0,1464	8,00E-08
rs1667394	15	26,20	A/G	G	0,1543	6,00E-09
rs8060502	16	88,12	A/G	G	0,3798	6,00E-09
rs4785698	16	88,20	C/G	G	0,1269	3,00E-09
rs164741	16	88,22	A/G	A	0,3277	4,00E-08
rs12918773	16	88,27	A/G	A	0,1042	7,00E-16
rs12922197	16	88,27	C/G	G	0,1047	4,00E-16
rs258322	16	88,28	C/T	T	0,1101	2,00E-15
rs7204478	16	88,32	A/G	A	0,4605	5,00E-07
rs7189734	16	88,33	A/G	A	0,26	3,00E-07
rs1805007	16	88,51	A/G	A	0,0946	8,00E-20
rs8049897	16	88,55	A/G	A	0,1622	2,00E-15
rs4785751	16	88,56	A/G	A	0,4895	7,00E-12
rs4785763	16	88,59	A/C	A	0,3548	2,00E-14
rs8057672	16	88,61	A/G	G	0,3099	9,00E-09
rs11648785	16	88,61	C/T	T	0,299	1,00E-08
rs868372	16	88,63	C/G	G	0,2914	2,00E-07
rs4911383	20	32,02	A/G	A	0,3434	8,00E-08
rs6059662	20	32,14	A/G	A	0,3381	2,00E-07
rs910873	20	32,64	A/G	A	0,1042	2,00E-10
rs17305573	20	32,64	A/G	G	0,1043	2,00E-10
rs4911442	20	32,82	A/G	G	0,1428	2,00E-07
rs2425025	20	33,31	A/G	G	0,0924	6,00E-10
rs17421899	20	33,40	C/G	G	0,1159	2,00E-07
rs2425067	20	33,67	A/G	G	0,1208	4,00E-07

Table S4. Transcription analyses of the 22 genes at 20q11.22 in 6 human skin melanocytic cell lines

	Averaged expression values (relative to ACTB)						
	LP22	**LP89**	**MP01**	**DP74**	**DP80**	**DP83**	**P-value**
RALY	0,0412	0,032	0,046	0,0476	0,0434	0,0309	0,4249
EIF2S2	0,0018	0,001	0,0112	0,006	0,0118	0,0017	0,0155
ASIP	nd	nd	nd	nd	nd	nd	nd
AHCY	0,0518	0,0363	0,0612	0,0462	0,0413	0,0423	0,5491
ITCH	0,0019	0,0011	0,0072	0,0034	0,0054	0,0023	0,0037
DYNLRB1	0,0797	0,0595	0,089	0,0997	0,1218	0,0307	0,3916
MAP1LC3A	0,017	0,0082	0,0186	0,0128	0,0083	0,0076	0,7739
PIGU	0,023	0,0151	0,0181	0,0245	0,0124	0,017	0,9961
TP53INP2	0,0048	0,0028	0,0084	0,0051	0,0031	0,0031	0,303
NCOA6	0,0027	0,0022	0,0066	0,0038	0,0057	0,003	0,0063
GGT7	0,0066	0,0028	0,0056	0,0057	0,0035	0,0042	0,739
ACSS2	0,0057	0,0036	0,0062	0,0046	0,0059	0,0056	0,0775
GSS	0,0117	0,0105	0,0165	0,0182	0,0132	0,0106	0,0942
MYH7B	nd	nd	nd	nd	nd	nd	nd
TRPC4AP	0,0208	0,0126	0,0264	0,023	0,0206	0,0182	0,0524
EDEM2	0,0076	0,0077	0,0107	0,0137	0,0108	0,0076	0,0356
PROCR	0,0016	0,0033	0,0054	0,0069	0,0057	0,0013	0,0645
MMP24	nd	nd	nd	nd	nd	nd	nd
MMP24AS1	0,0086	0,0048	0,0073	0,0084	0,0057	0,0069	0,8583
EIF6	0,035	0,0262	0,0466	0,0478	0,0332	0,022	0,4675
FAM83C	nd	nd	nd	nd	nd	nd	nd
UQCC	0,0016	0,001	0,0026	0,0025	0,002	0,001	0,1561

Significant associations are highlighted in colors (green P<0.05; yellow P<0.01; pink P<0.001)

Table S5. Transcription analyses of the 22 genes at 20q11.22 in 29 epidermal samples (light n=17; dark n=12).

	Average expression (relative to *ACTB*)		
	Light	**Dark**	**P-value**
RALY	0,075	0,097	0,00913
EIF2S2	0,067	0,099	0,00263
ASIP	nd	nd	nd
AHCY	0,048	0,061	0,02692
ITCH	0,042	0,053	0,00689
DYNLRB1	0,323	0,378	0,11068
MAP1LC3A	0,098	0,098	0,9877
PIGU	0,023	0,029	0,05065
TP53INP2	0,001	0,001	0,53808
NCOA6	0,042	0,053	0,0106
GGT7	0,005	0,006	0,12008
ACSS2	0,026	0,028	0,56577
GSS	0,022	0,032	0,00003
MYH7B	nd	nd	nd
TRPC4AP	0,037	0,044	0,10646
EDEM2	0,01	0,013	0,03147
PROCR	0,004	0,007	0,0318
MMP24	nd	nd	nd
MMP24AS1	0,007	0,007	0,84633
EIF6	0,211	0,254	0,05332
FAM83C	nd	nd	nd
UQCC	0,004	0,005	0,06519

Significant associations are highlighted in colors (green P<0.05; yellow P<0.01; pink P<0.001)

Table S6. Association between the transcription levels of the 22 genes at 20q11.22 and the rs1885120 genotypes in 6 human skin melanocytic cell lines

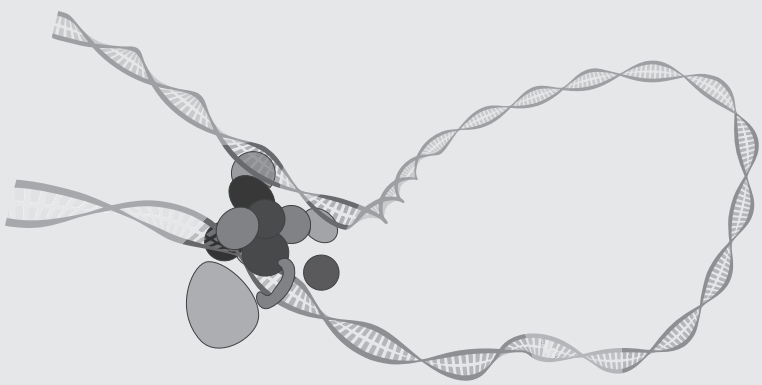
	rs1885120		
	CC	**GC**	**P-value**
RALY	0,037	0,047	0,105
EIF2S2	0,004	0,009	0,339
ASIP	nd	nd	nd
AHCY	0,043	0,054	0,181
ITCH	0,003	0,005	0,219
DYNLRB1	0,073	0,094	0,499
MAP1LC3A	0,01	0,016	0,229
PIGU	0,017	0,021	0,318
TP53INP2	0,003	0,007	0,056
NCOA6	0,003	0,005	0,295
GGT7	0,004	0,006	0,335
ACSS2	0,005	0,005	0,815
GSS	0,011	0,017	0,005
MYH7B	nd	nd	nd
TRPC4AP	0,018	0,025	0,096
EDEM2	0,008	0,012	0,07
PROCR	0,003	0,006	0,111
MMP24	nd	nd	nd
MMP24AS1	0,007	0,008	0,35
EIF6	0,029	0,047	0,017
FAM83C	nd	nd	nd
UQCC	0,001	0,003	0,032

Significant associations are highlighted in colors (green P<0.05; yellow P<0.01; pink P<0.001).

Table S2. *(available on request)*

Chapter 6

Discussion



Parts of this chapter were published in:

Genetic variation in regulatory DNA elements: the case of *OCA2* transcriptional regulation.

Mijke Visser, Manfred Kayser, Frank Grosveld and Robert-Jan Palstra

Pigment Cell & Melanoma Research, 27, 169-177 (2013)

With the emergence of large-scale genetic association studies like GWASs, numerous SNPs have been associated with a phenotype of interest, and it can be expected that many more will follow. The biology behind these genetic associations were initially assumed to be rather straightforward; effecting protein function, because at the time it was thought that causal SNPs were mostly protein coding. However, it turned out that the vast majority of the GWAS tag SNPs are located in intergenic or intronic regions. Determining the functional biology of these noncoding SNPs is much more of a challenge. Pigmentation-related traits have been extensively studied by genetic studies like linkage analysis, cytogenetics and candidate gene studies[1–3], and more recently in several GWASs[4], resulting in the identification of novel DNA variants highly associated with human pigmentation traits. In this thesis, I investigated the functional biology behind several pigmentation-associated SNPs identified by genome-wide and candidate-gene association studies.

Transcriptional regulation of pigmentation genes

In the work summarized in this thesis it has been proven highly valuable for the functional interpretation of pigmentation-associated SNPs to apply combinations of *in silico* analyses and various molecular experiments. Our functional studies were initiated with SNPs that were shown to be strongly associated with pigmentation phenotypes in previous GWASs (**Chapter 2-4**) and in novel GWASs on skin pigmentation (**Chapter 5**).

The SNP rs12913832 located in intron 86 of *HERC2* is a near perfect predictor of human eye color[5–7]. The intronic SNPs rs10756819 and rs2153271 are located in the first intron of the *BNC2* gene and are strongly associated with skin pigmentation[8] and freckling[9], respectively. Another intronic SNP rs12203592, located in the *IRF4* gene, is strongly associated with human pigmentation in general, including hair, eye and skin pigmentation, tanning response and nevus count[8–12]. The GWASs described in **Chapter 5** confirm the significant associations with skin pigmentation of these four SNPs, and in addition a SNP in the region of the known pigmentation gene *ASIP*, rs6059655 was identified to be strongly associated with quantitative skin color saturation.

The functional genetic studies presented in this thesis resulted in detailed, allele-specific characterizations of three enhancer elements in which the identified causal SNPs reside, and these functional SNPs modulate transcriptional regulation of three pigmentation target-genes in a highly tissue-specific manner.

Transcriptional regulation of pigmentation genes – The case of OCA2

Transcription of the pigmentation gene *OCA2* is regulated by an intronic enhancer element located in the neighboring gene *HERC2* in skin melanocytes. The SNP rs12913832 is located in this enhancer and is predicted to alter a DNA binding site for the transcription factor HLTF. Furthermore, several DNA binding sites for the transcription factors MITF and LEF1 are also predicted to be present within this region[6]. Both MITF (microphthalmia-associated transcription factor) and LEF1 (lymphoid enhancer factor 1) are crucial in melanocyte development[13,14]. As a DNA-dependent chromatin-remodeling ATPase, HLTF is involved in the repositioning of nucleosomes to allow access of transcription factors to their related

binding sites[15]. Together with my collaborators, I demonstrated robust binding of HLTF, LEF1 and MITF at the rs12913832-enhancer region when the rs12913832 T-allele is present, whereas with the C-allele, binding of these transcription factors was not detected at the region around rs12913832, confirming and refining the proposed model of Sturm et al[6] (and at the same time refuting the proposed model of Eiberg et al[5]). Since HLTF functions as a chromatin remodeler, it has been suggested that preferential interaction of HLTF with the rs12913832 T-allele facilitates binding of MITF and LEF1 by allowing access to the chromatin[6]. Additionally, LEF1 was previously shown to interact with MITF[16] and to be involved in chromatin looping[17]. We indeed demonstrated a long-range chromatin loop between the rs12913832 enhancer and the *OCA2* promoter that is more prominent when the rs12913832 T-allele is present (Figure 1), which is probably due to increased binding of the involved transcription factors.

Additionally, an allele-independent chromatin loop was detected from the rs12913832 enhancer as well as from the *OCA2* promoter towards a potential regulatory element in the intergenic region between *HERC2* and *OCA2*. Analysis of online available datasets for accessibility and transcription factor binding in melanocytes, as well as our ChIP-seq data for acetylated histone 3 (H3K27Ac) (Palstra et al, manuscript in preparation) suggests the presence of this (and possibly also other) potential regulatory elements within the *OCA2-HERC2* locus (Figure 2).

Many noncoding DNA variants within the *OCA2-HERC2* region have been associated

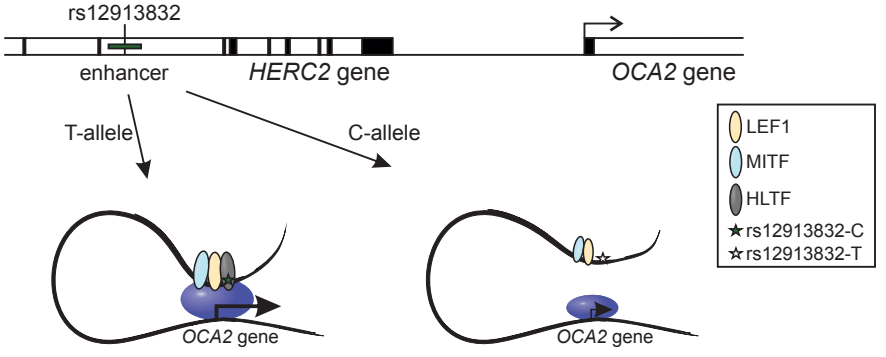


Figure 1. Model representing how sequence variation in the distal enhancer element in *HERC2* affects transcription of *OCA2*. The pigmentation-associated SNP rs12913832 is located in an enhancer element in intron 86 of *HERC2*. With the rs12913832 T-allele present, the transcription factor HLTF (grey oval) binds to the enhancer, which allows for the recruitment of MITF (light blue oval) and LEF1 (yellow oval), chromatin loop formation, and proper activation of *OCA2*, and this results in the dark pigmentation phenotype. In contrast, with the rs12913832 C-allele present, the enhancer is unable to bind HLTF, which leads to reduced recruitment of MITF and LEF1 and decreased *OCA2* expression, and this results in the light pigmentation phenotype. Larger dark blue oval: transcription complex, black arrows: level of transcription.

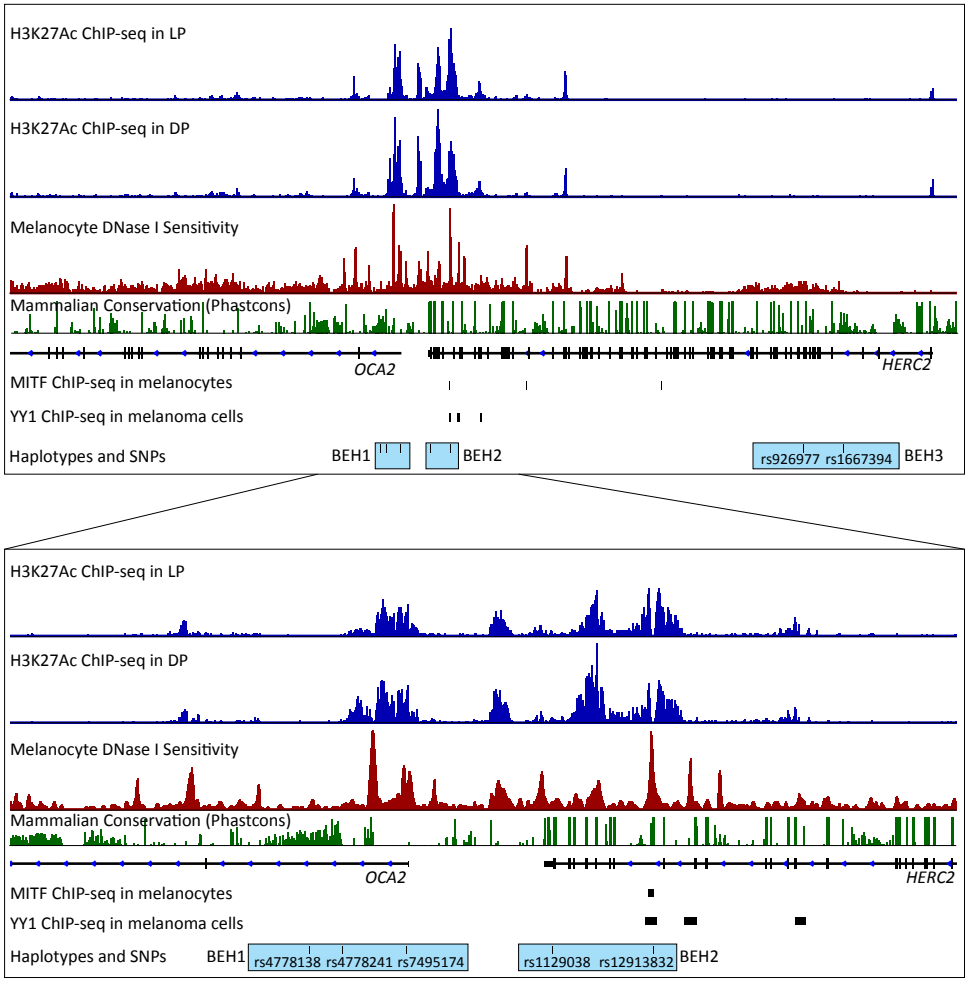


Figure 2. Overview of the *OCA2-HERC2* locus. IGV genome browser[18] shows a 400 kb window (top panel) and 80 kb zoom in (bottom panel) of the *OCA2-HERC2* locus in the 15q11-q13 chromosomal region. ChIP-seq data sets for acetylated histone 3 (H3K27Ac) (Palstra et al, manuscript in preparation) in light (LP) and dark (DP) pigmented melanocytes are shown in blue. Melanocyte open chromatin by DNaseI HS from ENCODE[19] is shown in red. Vertebrate conservation by PhastCons is shown in green. Binding signals of transcription factors MITF[20] and YY1[21] are depicted by black bars. The approximate position of the blue-eye-associated haplotypes (BEH)[22] are indicated as light blue boxes, with the SNPs that comprise the BEHs depicted in the respective boxes for BEH1, BEH2 and BEH3.

with pigmentation phenotypes at different levels of significance. Most of these DNA variants fall into three different haplotype blocks; the so-called blue-eye haplotypes (BEHs)[22]. BEH1 consists of three SNPs located in the 5' region of the *OCA2* gene[23], the second haploblock (BEH2) is located at the 3' region of *HERC2*[6], and it includes the rs12913832 enhancer and the SNP rs1129038, which is in nearly complete linkage disequilibrium (LD) with rs12913832.

BEH3 consists of SNPs rs916977 and rs1667394 and it is located more proximal to the *HERC2* promoter region[24,25]. Chromatin profiling indicates that both BEH1 and BEH2 contain robust peaks of DNaseI HS (DHS), signals of the active chromatin mark H3K27Ac, and binding of MITF and YY1 in BEH2 (Figure 2), suggesting regulatory potential. In contrast, BEH3 is located in a region of low DNase sensitivity, reduced H3K27Ac enrichment and absent TF binding, suggesting that this region has limited regulatory capacity. Despite the regulatory potential of the BEH1 and BEH2 regions, no evidence was found for the three SNPs that comprise BEH1 and rs1129038 being part neither of BEH2 nor for their LD SNPs (R^2>0.8) to be involved in (transcriptional) regulation such as rs12913832. It is therefore highly unlikely that any of the SNPs other than rs12913832 that belong to the *OCA2-HERC2* haplotype is involved in the allele-specific transcriptional regulation of *OCA2*; however, this does not exclude the possibility for a SNP from a different, more distant, haplotype to additionally affect *OCA2* expression.

Recently, a GWAS was performed on eye and skin pigmentation in individuals from Cape Verde, where extensive West African/European admixture resulted into a wide range of variation in both eye and skin color[26]. The SNP rs4424881 located in intron 1 of the *APBA2* gene was identified to be strongly associated with skin color, this gene has no known function in human pigmentation and is located about 1 Mb away from the *OCA2* gene. Despite of being in the haplotype radius of 1 Mb, rs4424881 belongs to a haplotype set that is distinct from the *OCA2-HERC2* (eye-color) locus. Beleza and colleagues hypothesize that (an LD SNP of) rs4424881 modulates an enhancer element that might be involved in the transcriptional regulation of *OCA2* affecting skin color specifically, while the rs12913832 enhancer is considered to control *OCA2* expression to affect eye color[26]. Using the IGV browser[18], we looked at the chromatin profile of the region surrounding the haplotype block of rs4424881, including DHS signals, H3K27Ac enrichment, and transcription factor binding in melanocytes, and this indeed implicates that at least one of the SNPs in high LD with rs4424881 (R2>0.8) is located in a potential enhancer element (data not shown).

The differences in *OCA2* expression between skin melanocytes with the rs12913832 C-allele and skin melanocytes with the T-allele are relatively modest (**Chapter 2**, Figure 1C, and Cook et al[27]), as well as the observed differences for the enhancer activity (**Chapter 2**, Figure 2E) and melanin content[27]. *OCA2* expression levels in iris melanocytes are unknown, but it has been reported that blue irises have minimal melanin content, whereas dark irises contain substantially larger amounts of both eumelanin and pheomelanin[28–30]. Since rs12913832 is very strongly correlated with eye color, this suggests that *OCA2* expression levels might differ extensively between blue and dark irises, more so than between light and dark melanocytes. It is therefore possible that the chromatin environment is more restricted in iris melanocytes as compared to that in skin melanocytes, driving the rs12913832 enhancer to act as an on-off switch instead of a rheostat. Nevertheless, this leaves open the possibility of other, more distant enhancers such as the potential regulatory regions in *APBA2*, to be involved in controlling *OCA2* expression in skin melanocytes. The hypothesis of such a pigment-tissue specific transcriptional regulation of *OCA2* could help explain the correlation and evolutionary history of skin and eye color, and supports the possibility that blue eyes were selected independently (and possibly for other reasons) from fair skin[26]. To test this hypothesis, future studies should investigate not only the regulatory potential of

the enhancer in *ABPA2* in skin melanocytes, but should use both iris and skin melanocytes to study tissue-specific patterns of *OCA2* expression levels, *OCA2* transcriptional regulation, and activity of the putative enhancers either within the *OCA2-HERC2* BEHs or more distant in *APBA2*.

Transcriptional regulation of pigmentation genes – The case of BNC2

The *BNC2* study described in **Chapter 3**, is a very appropriate example of how an associated noncoding tag SNP identified by a candidate-gene approach, leads to the discovery of the actual causal SNP by studying functional genetics of the tag SNP as well as its LD SNPs in combination with *in silico* regulatory datasets. Following the finding of the functional SNP rs12350739 in the case of *BNC2*, the region around this SNP was characterized as an enhancer element involved in the allele-specific transcriptional regulation of *BNC2* in skin melanocytes. Together with an alternative promoter at exon 3 and a potential regulatory element in intron 2, a model is proposed for the melanocyte-specific and allele-dependent transcriptional regulation of *BNC2* (Figure 3). With the G-allele present, the chromatin region around rs12350739 is open and more enriched for H3K27Ac, the activity of the enhancer is relatively high, resulting into increased *BNC2* expression (Figure 3A). In contrast, with the C-allele, the chromatin is closed and not enriched for H3K27Ac and the enhancer is only moderately active, resulting into decreased *BNC2* expression (Figure 3A). Based on DNase sensitivity data obtained from ENCODE[19] in a broad variety of tissues and cell lines, the combination of DHS signals at the rs12350739 enhancer and the (potential) intronic regulatory element seems to be highly specific for melanocytes and melanoma cells (Figure 4). Since the majority of the *BNC2* transcripts originate from the alternative rather than from the constitutive promoter, it is likely that the intergenic rs12350739 enhancer together with the (potential) regulatory element in intron 2 controls transcription of *BNC2* by physically interacting with the alternative promoter via a chromatin loop (Figure 3B), which eventually leads to a differentially expressed isoform that has a specific function in pigmentation pathways.

The finding of an alternative promoter in skin melanocytes and potentially more than one (tissue-specific) enhancer affecting this promoter is not fully unexpected, as it was already shown that *BNC2* has (at least) six promoters and many more splicing isoforms resulting into more than 90.000 possible mRNA isoforms encoding more than 2000 different proteins[31]. However, even though *BNC2* is abundant in a variety of tissues and cell types[32,33], its expression and subsequently its cellular functioning is most probably regulated in a dedicated and highly tissue-specific manner, which also reflects the biological function of *BNC2* that was suggested to be essential, due to its very high conservation status[32,34,35].

Notably, several SNPs falling into the same LD block as rs12350739 are associated with ovarian cancer susceptibility and with ovarian abnormality[36,37]. Consequently, and due to the presence of comparable DHS signals at the region around rs12350739 in myometrial cells[19], we consider it possible that the rs12350739 enhancer is also active in uterine and ovarian tissue. However, the robust DHS signals detected in skin melanocytes at the regulatory element in intron 2 are not observed in myometrial cells, which implies that this potential regulatory element is not involved in the transcriptional regulation of *BNC2* in these

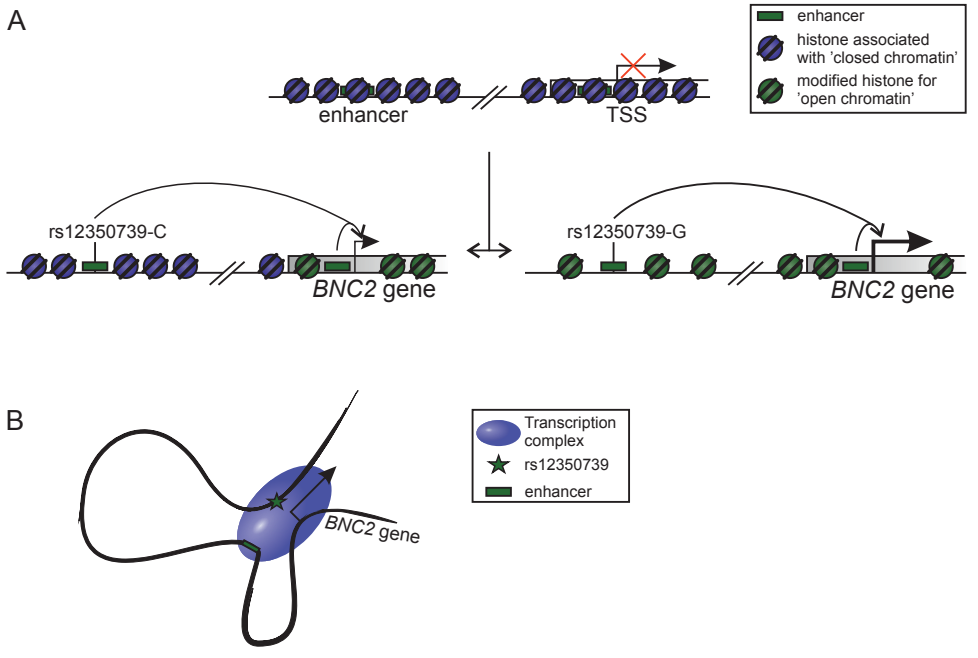


Figure 3. Model depicting how expression of *BNC2* is influenced by a DNA variant located in an intergenic enhancer element together with an alternative promoter and an intronic potential regulatory element. (A) the actual causal SNP rs12350739 that is in LD with the pigmentation-associated intronic SNP rs10756819 is located in a tissue-specific enhancer element 14 kb upstream from the constitutive promoter of *BNC2*. Furthermore, an alternative promoter and a tissue-specific potential regulatory element are present at exon 3 and in intron 2, respectively. When the rs12350739 G-allele is present, the chromatin of the region around rs12350739 is accessible and enriched for H3K27Ac, the rs12350739-enhancer is active, resulting into increased *BNC2* expression. In contrast, with the rs12350739 C-allele present, the chromatin is hardly accessible and not enriched for H3K27Ac, resulting into a moderately active enhancer and diminished expression of *BNC2*. (B) As the majority of *BNC2* transcripts originate from the alternative promoter at exon 3, I propose that the rs12350739-enhancer cooperates with the potential tissue-specific regulatory element in intron 2 to activate *BNC2* transcription, which can be achieved by forming long-range loops from both enhancers to the alternative promoter, ultimately resulting into a differentially-expressed, tissue-specific isoform of *BNC2*. Black arrows: level of transcription.

cells, and we speculate that this results into yet another combination of *BNC2* isoforms and expression levels in this particular cell-type. This serves as another illustration of the highly dedicated and tissue-specific manner genes that can be transcriptionally regulated.

Transcriptional regulation of pigmentation genes – The case of IRF4

In contrast with the transcriptional regulation of *OCA2* and *BNC2* as described in the previous paragraphs, the responsible allele-specific enhancer element that activates *IRF4* expression is located in an intron of the *IRF4* gene itself, rather than outside the gene in the intergenic

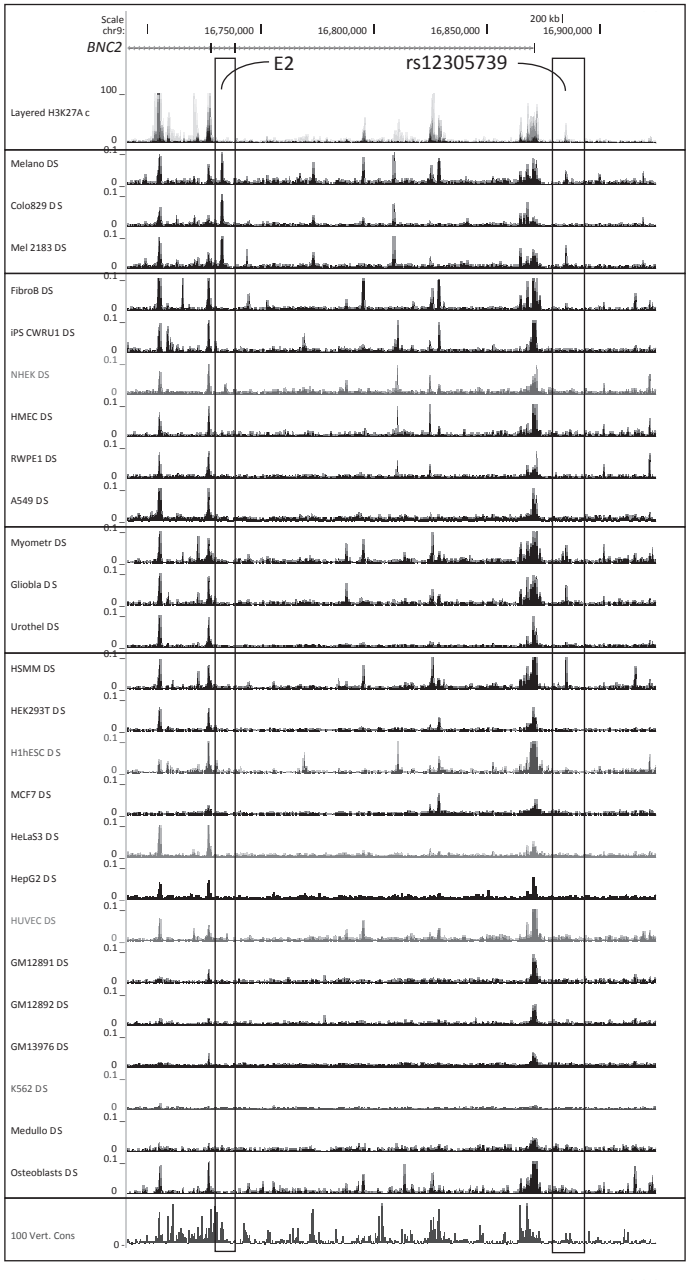


Figure 4. The combination of DNaseI HS signals at the rs12350739-enhancer and the potential regulatory element in intron 2 is highly specific for melanocytes and melanoma cells. Tracks from the UCSC browser displaying DNaseI HS (DHS) signals at the different regulatory regions of *BNC2* in a broad range of tissues and cell-types. Note that DHS signals at the rs12350739-enhancer (boxed and indicated with rs12350739) are also present in myometrial, glioblastoma and HSMM (skeletal muscle myoblasts) cells, however not in combination with DHS signals at the potential regulatory element in intron 2 (boxed and indicated with E2).

region (i.e. *BNC2*) or in an intron of a neighboring gene (i.e. *OCA2*). In **Chapter 4** I describe the characterization of this enhancer and the subsequent transcriptional regulation of *IRF4*, and I demonstrate that both are highly tissue-specific as well as strongly dependent on the allelic status of the pigmentation-associated intronic SNP rs12203592. Besides using

conventional techniques for the detection of histone mark H3K27Ac, chromatin accessibility, transcription levels and chromatin loop formation, together with my collaborators I used a new experimental approach to specifically measure the allelic contribution of rs12203592 to the functional characteristics of the enhancer. This allele-specific approach uses a heterozygote cell line as test tube in which the cellular environment is stabilized for both tested alleles. This excludes the possibility of (unknown) trans-acting factors modulating the transcriptional regulation of the gene instead of the allelic differences of the SNP of interest. Based on the data obtained in our study, as well as on previously published studies, we proposed a model for the transcriptional regulation of *IRF4* in melanocytes (graphically displayed in **Chapter 4**, figure 7). With the rs12203592 C-allele present, the transcription factor TFAP2α binds to the enhancer in intron 4[38], which allows for the recruitment of the transcription factors MITF, YY1, and presumably LEF1, chromatin interactions with the *IRF4* promoter and potentially with an additional regulatory element in intron 7, resulting into proper and stable transcription of *IRF4*. With the rs12203592 T-allele, TFAP2α is unable to bind[38], which leads to reduced recruitment of MITF, YY1 and LEF1, less stable chromatin interactions and consequently, decreased *IRF4* expression.

In this study on *IRF4*, as well as in the studies previously described on *BNC2* and *OCA2*, additional potential regulatory elements were identified that are likely to be involved in the transcriptional regulation of these genes. These findings are not surprising, as already more than a million putative enhancers were identified in the human genome[39,40], which outnumbers the amount of genes of approximately 20.000[41] tremendously, and suggests that even with a large number of enhancers being poised, still multiple enhancers, both constitutive as well as tissue specific are involved in the transcriptional regulation of every gene[42].

Transcriptional regulation of pigmentation genes – The case of the 20q11.22 region

Although the studies described in **Chapters 2, 3 and 4** exemplify that the functional characterization of association signals can be relatively straightforward, the study described in **Chapter 5** however demonstrates that this can also be more complex. By combining a series of GWASs in more than 17.000 Europeans, five genomic regions were identified harboring SNPs that are significantly associated with skin color phenotypes. The biology of the genes and SNPs located in four of these regions have been described already previously, of which two were also described in this thesis (*OCA2* in **Chapter 2** and *IRF4* in **Chapter 4**), we therefore aimed to characterize the association signals that were identified in the fifth region; 20q11.22. The top-associated SNP rs6059655 is located in intron 8 of the *RALY* gene, in a haplotype block of ~1.5 Mb containing 22 known genes, including *ASIP*, a gene encoding the agouti signaling protein that is known to be involved in pigmentation biology. Chromatin profiling of the 20q11.22 region using *in silico* regulatory data sets resulted into the identification of numerous potential regulatory elements in this haplotype block. However, none of these elements coincided with the physical locations of any of the associated tag SNPs or their LD SNPs. Expression analyses of the 22 genes in the 20q11.22 region using epidermal cells highlighted two novel genes, *EIF2S2* and *GSS*, as the most-likely candidates responsible for the observed genetic associations. Analysis of eQTL data

in full-layer skin biopsy material, using the online available tool GeneVar revealed a strong correlation between the expression of *ASIP* and the genotypes of three associated SNPs. Remarkably, expression of this gene was detected only in the skin dermis, and not in the epidermal layers of skin in which the majority of the pigmentation biology takes place. This implies that the observed association signals might also exert their function in dermal cells rather than in epidermal cells.

The exact functional mechanisms that link the associated SNPs to the transcription of the two highlighted genes in epidermal tissue as well as to *ASIP* in dermal tissue remains to be elucidated. Several scenarios are open for further investigation, such as the functional SNP could affect DNA folding, posttranslational control mediated by miRNAs or transcriptional regulation mediated by lncRNAs or by enhancer elements in which yet to be discovered associated (LD-)SNPs reside. Since the 20q11.22 region is particularly large, it is most likely that more than one SNP is functional and it is crucial for all plausible scenarios to identify the actual causal SNP(s). This may be achieved by investigating the chromatin profile using additional *in silico* regulatory datasets suitable for the different scenarios, by including deeper sequencing of transcripts[43] derived from the targeted 20q11.22 region, to identify rare transcripts such as lncRNAs, by using a variety of computational tools that for example predict miRNA binding sites and by including more cell lines and tissues to avoid false positive and negative results obtained due to tissue-specific effects.

Involvement of the studied (new) pigmentation genes in melanogenesis

The elucidation of the functional biology underlying an association signal not only provides insights into the regulatory function of the causal SNP and the enhancer it (potentially) modulates, it might also provide novel insights into the biological role of the target gene. Although it was not the scope of this thesis, I briefly consider the (potential) biological roles of the main target genes in light of the studied phenotype. In **Chapter 2** I describe the allele-specific transcriptional regulation of *OCA2*. As a known pigmentation gene, *OCA2* plays an important role in melanogenesis and controls the eumelanin content in melanocytes[44]. Its gene product, the P-protein is suggested to be involved in small molecule transport across the melanosomal membrane[45], regulating melanosomal pH, which in turn affects processing, trafficking and activity of TYR[46,47].

The allele-specific and tissue-dependent transcriptional regulation of *BNC2* is described in **Chapter 3**. This gene is expressed in many tissues and due to its high evolutionary conservation its function is thought to be essential. It is most likely involved in mRNA processing[32,33], but it has also been suggested to function as a transcription factor[34]. Together with my collaborators I demonstrated that *BNC2* is differentially expressed in melanocytes, potentially in the form of a melanocyte-specific isoform and it has been proposed that *BNC2* influences expression of the pigmentation gene *KITLG* in melanocytes[48].

Chapter 4 describes how an allele-dependent intronic enhancer modulates expression of the *IRF4* gene. As a member of the *IRF* family, *IRF4* functions as a DNA-binding transcription factor involved in immune system development and response, but it most-likely also has a role in pigmentation biology or melanocyte development. Recently it was reported that expression of *TYR* is regulated by cooperation between *MITF* and *IRF4*[38].

Of the three genes highlighted at the 20q11.22 region in **Chapter 5**, *ASIP* is most well-known for its involvement in pigmentation biology, it functions as an antagonist for the G-coupled receptor MC1R and as such it is involved in the switching between eumelanogenesis and pheomelanogenesis[49,50]. Expression of *ASIP*, however, was not detected in cultured melanocytes (**Chapter 5** of this thesis, and recently shown by Haltaufderhyde et al[51]), or in the skin epidermis which consists of melanocytes and keratinocytes, while in the skin dermis, expression of *ASIP* was robustly detected (**Chapter 5** of this thesis). It is known that besides the skin epidermis, also components from the skin dermis contribute to normal skin pigmentation, such as dermal fibroblasts[52], and melanocyte stem cells (MelSCs)[53]. Dermal fibroblast are known to be involved in the secretion of several (paracrine) factors that modulate signaling pathways involved in melanocyte function[52], it is therefore possible that ASIP, as an antagonist of MC1R, becomes secreted by these fibroblasts to interact with MC1R. It is however also possible that *ASIP* is expressed in other cells, like MelSCs where it might act on pigmentation in an earlier developmental stage and might become silenced upon differentiation of the melanocytes in the epidermis. These scenarios, and possibly others, remain to be studied in future studies in order to pinpoint the exact cell-type and/or development stage *ASIP* is expressed in to exert its function in pigmentation biology. Furthermore, these results also stress the importance of using the proper cell-type or tissue to study the functional biology behind an association signal and the related gene.

Glutathione synthetase encoded by the *GSS* gene, is involved in the catalyzation of the second step of the glutathione (GSH) biosynthesis and it was also highlighted at the 20q11.22 region. This highly important cellular antioxidant has multiple cellular functions and it exerts major effects on melanogenesis in melanocytes[54–57]. It was shown to play an important role in the switching between eumelanogenesis and pheomelanogenesis through interaction with TYR[54], and by reacting with dopaquinone in the TYR pathway[55,56]. Furthermore, the pigmentation gene *OCA2* was suggested to be involved in controlling metabolism of GSH[58].

In contrast to *ASIP* and *GSS*, the role in pigmentation biology of the third prioritized gene at 20q11.22, *EIF2S2* remains to be unraveled. The only indirect evidence for such a role was provided in studies on the agouti-yellow (A^y) deletion in mice. The A^y mutation deletes *Eif2s2* together with neighboring gene *Raly*, and causes ectopic expression of agouti, resulting into several phenotypes, including yellow fur[59,60]. *EIF2S2* encodes the beta subunit of the eukaryotic initiation factor (eIF2), which plays a central role in translation initiation. During the early steps of protein synthesis, eIF2 forms a ternary complex with GTP and methionine-loaded initiator tRNA. This complex binds other initiation factors and ribosomal subunits, to form a pre-initiation complex that binds and scans mRNA, recruits more initiation factors and ribosomal subunits ultimately resulting into initiation of translation[61]. The beta subunit of eIF2 (eIF2beta) catalyzes the exchange of GDP for GTP, which recycles the eIF2 complex for another round of initiation[62]. Besides the described phenotypes of the A^y deletion in mice, deletion of eIF2beta was reported to suppress testicular cancer incidence and reduce lethality in yellow agouti mice[63]; however, a (potentially specific translation-initiation) role in pigmentation was not yet reported, and remains to be investigated in future studies.

In figure 5 I graphically present a simplified combination of signaling pathways including the main well-known players involved in melanogenesis, and the genes highlighted

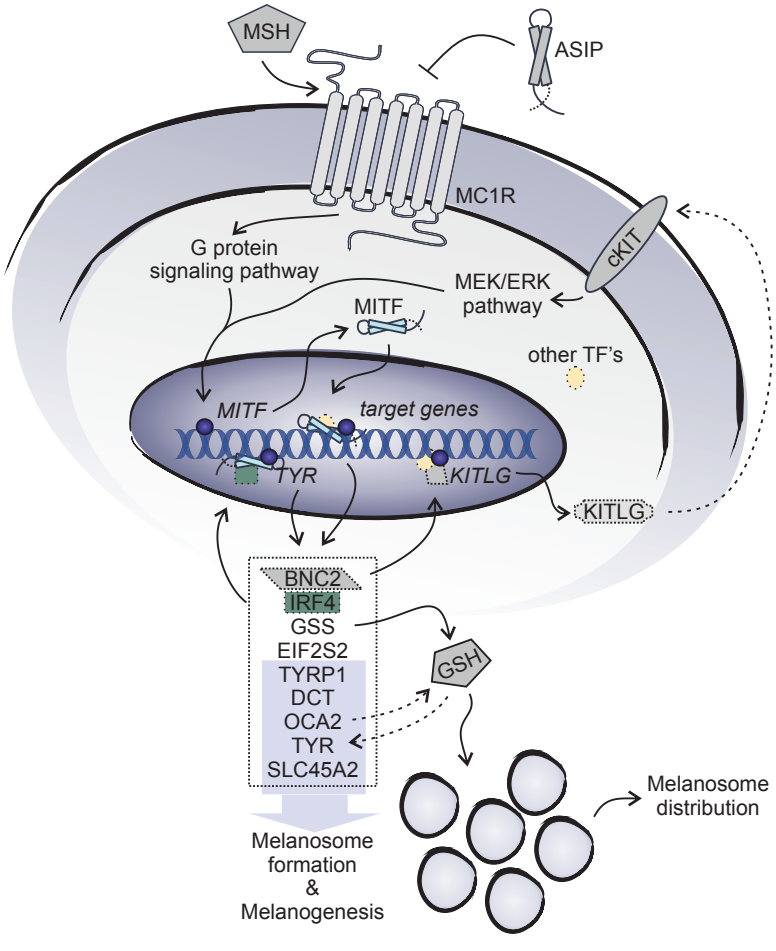


Figure 5. Roles of *OCA2*, *BNC2*, *IRF4*, *ASIP*, *GSS* and *EIF2S2* in pigmentation biology. Based on well-known signaling events involved in melanogenesis, as displayed in the introduction of this thesis (**Chapter 1**, figure 4), the involvement of the genes studied in the chapters of this thesis are presented. Binding of melanocyte-stimulating hormone (MSH) activates the trans-membrane MC1R receptor, resulting into a cascade of signaling events; activated MC1R stimulates expression of the transcription factor *MITF* via G protein signaling. As the master regulator in melanocytes, MITF controls transcription of several genes involved in melanogenesis, including *TYR*, *TYRP1*, *DCT*, *SLC45A2*, *OCA2*, *BNC2*, *IRF4*, and presumably *GSS* and *EIF2S2*. In turn, *BNC2* influences expression of the pigmentation gene *KITLG*[48], *IRF4* cooperates with MITF to regulate transcription of *TYR*[38], and GSS is involved in biosynthesis of glutathione (GSH), a key component in the eu/pheomelanogenesis swith[54,56]. GSH displays a dual effect on TYR, depending on the cellular concentration of GSH. Besides its central role in melanosome biogenesis, *OCA2* was also suggested to be involved in controlling GSH metabolism[58]. MC1R activation can be inhibited by ASIP, resulting into decreased activation of the target genes of MITF, which eventually leads to a switch to pheomelanogenesis. Adapted from Law et al[64].

in **chapters 2-5** of this thesis. The actual mechanism is much more complex, and I anticipate that many more factors, genes and genetic variations will be identified to affect the various stages and processes of pigmentation biology.

From GWAS to function

Based on the functional studies described in this thesis, recently published reports, and online available data and bioinformatics tools, I will discuss in the coming sections an approach that can be used as a general research strategy to study the functional biology behind genetic association signals (Figure 6).

Identification of the putatively causal DNA variant

GWASs have identified more than 11.000 SNPs that are associated with a large variety of human traits and disease phenotypes[65]. Due to the original design of GWASs and the SNP arrays used herein, these SNPs are typically so-called 'tags' for haplotypes on which the directly functional or causal variants reside[66,67]. A causal variant is a SNP that influences a molecular or cellular process to affect a human phenotype. Only ~16% of phenotype-associated loci harbor SNPs that are protein-coding, the remaining SNPs are noncoding located in intronic or intergenic regions[68]. The first challenge in elucidating the functional biology behind these identified genetic association is therefore to pinpoint the actual causal DNA variant, hence the SNPs that are in (high) LD with the phenotype-associated SNP have to be included in functional analyses. These LD SNPs can be retrieved using several key resources, including the HapMap[69] and 1000 Genomes[70] projects. Bioinformatics tools such as Haploview[71] can be employed to visualize and manipulate downloaded genotype data covering specific loci, and the SNP Annotation and Proxy (SNAP)[72] search tools at the Broad Institute enables plotting of association data with LD information.

In silico annotation

It has been shown by numerous studies that many noncoding SNPs locate in regulatory elements and influence gene expression through transcriptional and posttranscriptional processes[74–81]. Several of these regulatory elements are characterized by a range of different histone modifications depending on the regulatory function of the element including histone acetylation and methylation, (highly) accessible chromatin, transcription factor binding, and long-range chromatin interactions (loop formation)[82]. Large-scale genomics projects, such as ENCODE[39], the NIH Roadmap Epigenomics Mapping Consortium[83], and the Functional Annotation of the Mammalian Genome (FANTOM)[84], as well as independent labs are exploring the human genome of numerous diverse cell and tissue types and across several developmental stages for regulatory characteristics. The data generated by these projects includes gene and isoform annotations, gene-expression levels analyzed with expression quantitative-trait loci (eQTL), genome-wide transcription factor binding and chromatin modifications analyzed with chromatin immunoprecipitation paired with massively parallel sequencing ChIP-seq, genome-wide chromatin-accessibility analyzed with FAIRE sequencing and DHS sequencing, long-range chromatin interactions analyzed

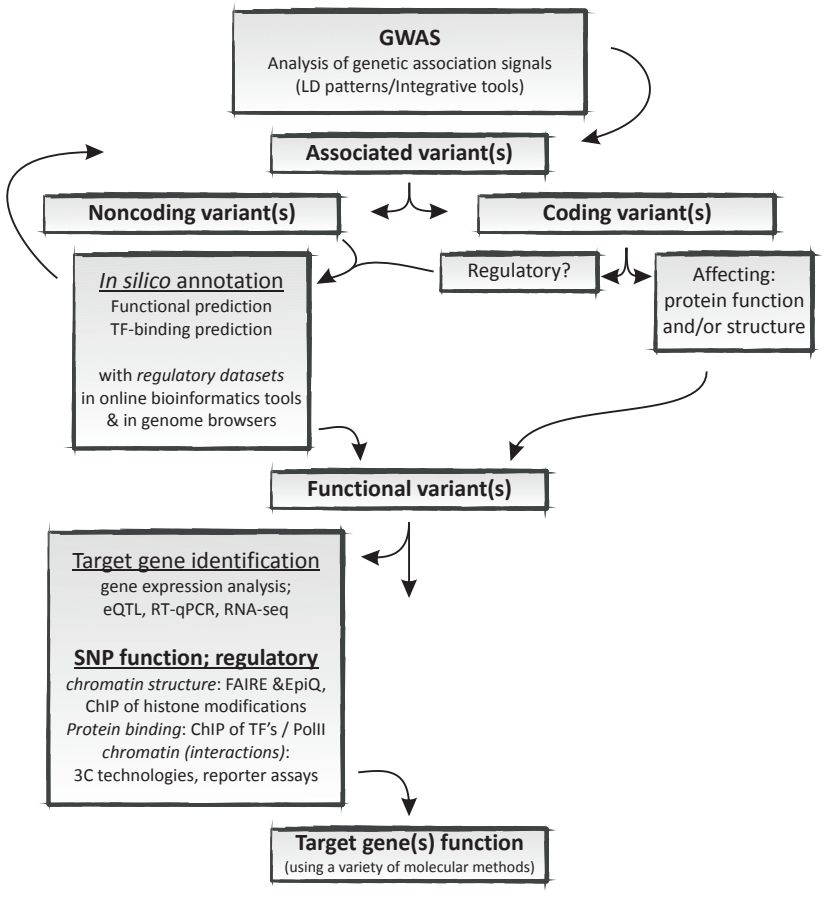


Figure 6. Strategy for functionally analyzing and interpreting genetic association signals. Starting with phenotype-associated DNA variants resulting from GWASs, the following steps should be taken to elucidate the functional biology underlying the association signal: 1. Identification of the putatively causal DNA variant. Include SNPs in linkage disequilibrium (LD) in the following analysis steps, 2. Analyzing in silico annotations to identify regulatory potential of the (LD-)SNPs, 3. Investigate gene expression patterns to identify the potential target gene(s), 4. Experimental confirmations of SNP function by employing a wide range of molecular methodologies and 5. Studying the biological function of the potential target gene(s). Details of the strategy steps can be found in the text.

with 3C-derived technologies, and evolutionary conservation, providing a broad variety of regulatory datasets.

As soon as an initial set of candidate SNPs including (high) LD SNPs is established, several computational (online-available) tools can be applied that specifically query these large-scale genome-wide datasets to screen the first, relatively large set of candidate variants for functional characteristics. Haplotype blocks are highly diverse in size, with an average length of 20-50 kb; however, some loci are >> 100 kb long, containing more than 1000 variants, while others are much smaller with only several variants that display

significant associations[85]. If the associated haplotype locus is not too long, the locus can be profiled by assessing the datasets without the use of a bioinformatics tool, but instead the datasets can be loaded into a genome browser like UCSC (http://genome.ucsc.edu/)[86] or IGV (www.broadinstitute.org/igv/)[18] for visual inspection. Mapping of epigenetic characteristics using ChIP-seq datasets, combined with accessibility data and evolutionary sequence conservation can identify potential regulatory elements in a region of interest. Additionally, several online available bioinformatics tools such as JASPAR[87] or TRANSFAC[88] can be used to computationally predict TF binding within such a region. Important to note is that with all applied *in silico* approaches, using bioinformatics tools and/or visual inspection, false positive and false negative results can be obtained. This is most probably due to the high tissue-specific characteristics of regulatory elements, resulting into the possibility to miss data leading to the absence of valid results (false negative) as well as the possibility of SNPs being predicted to influence regulatory signals in irrelevant tissues (false positive)[85]. This can be essentially avoided by carefully choosing regulatory datasets that are acquired in relevant cell-types or tissues.

Following the selection of candidate variants, a wide range of experimental approaches should be applied to confirm and if necessary further pinpoint the functional variant.

Towards identification of the target gene(s)

SNPs that disrupt a coding sequence of a gene can clearly implicate the corresponding phenotype-associated gene. However, identifying the target gene of noncoding variants can be more challenging. Expression of genes is highly heritable and regulated in a dedicated, and often tissue-specific manner. If a noncoding SNP is involved in the transcriptional regulation of a gene, the expression of that gene should correlate with the SNP-genotype, which should be investigated by measuring gene expression levels with Reverse-Transcription quantitative PCR (RT-qPCR) or (targeted) RNA sequencing, and linking these expression levels with the genotypes of the SNP(s) of interest. This approach can be applied when already existing functional evidence prioritizes a candidate gene (e.g. *OCA2*), or when the associated haplotype locus contains only a few genes. However, when the haplotype locus is relatively large containing numerous SNPs and genes, and the target gene(s) is (are) unknown, a less laborious approach, such as analyzing eQTLs, should first be applied to prioritize candidate target genes.

Genetic associations that correlate with gene expression are known as expression quantitative trait loci (eQTL)[85,89], and it was shown in multiple studies that GWAS signals are enriched with eQTL in a tissue-specific manner. Studying eQTL can therefore be very useful for (further) understanding genetic association signals, and for the identification and/or confirmation of (candidate) target genes. Several (online) resources such as GeneVar are now available to analyze eQTLs[90]. No prior knowledge of functional mechanisms is needed for the associations between alleles and target genes, as the eQTL signals are annotated in an unbiased manner. However, it is particularly crucial to study eQTL effects in appropriate tissue or cell types, as 50-90% of eQTL are estimated to be tissue-specific[90,91] and as already noted previously, the trait-associated variants are also known to exert tissue-specific

effects[92]. Besides tissue dependency, other influencing factors should additionally be considered when studying eQTL data, such as *cis* and *trans*-acting effects[93], composition of the haplotype[94], coding and noncoding RNAs[95], alternative splicing[96] and variation in mRNA stability[97]. Finally, experimental confirmation of potential links between genotype and gene transcription that were highlighted by eQTL analysis is highly recommended.

Experimental confirmations of the putative SNP function

Elucidation of the involved regulatory mechanism and verification of the obtained *in silico* and computational prediction data requires extensive experimental studies. As these regulatory mechanisms are often highly tissue-specific, the use of relevant cell lines and/or human or mouse tissue is crucial. Molecular methodologies such as ChIP-qPCR and FAIRE or EpiQ should be employed to study transcription factor binding, histone modifications and chromatin accessibility at the regions of interest. RT-qPCR analysis should be used to confirm eQTL data and/or prioritize target genes in (a larger sample set of) relevant cell lines or tissue. Besides confirmatory analyses, it is also highly informative to include allele-specific analyses of the regulatory characteristics by selecting suitable cell lines or tissue samples with different genotypes of the functional variant(s). Ideally, cell lines or tissue samples that are heterozygote for the functional SNP should be used to compare regulatory characteristics of the opposite alleles in the same cellular environment, which excludes the possibility of differential trans-acting factors instead of the allelic differences to modulate the detected variation in transcriptional regulation. Following (allele-specific) characterization of a potential enhancer element, its activity can be tested using reporter assays. In these assays, regulatory elements with either the minor allele or the common allele are cloned into a promoter-driven reporter construct and transiently transfected into relevant cell lines. Comparing the effects of the different alleles on the activity of the enhancer then provides further insight into the mode-of-action of the enhancer element.

Even more powerful are the more recently developed genome-editing technologies such as TALEN (transcription activator-like effector nuclease) and CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated genes)[98,99], that can be used to artificially generate a cell line or mouse model of genetic association variants by introducing mutations into the genome. In these model systems the regulatory mechanisms can be studied by employing molecular methodologies as described above in order to elucidate the functional biology that underlies the associated variant. Additionally, the CRISPR/Cas system allows mutagenesis at multiple loci in one cell, which provides the possibility to study the joint effects of multiple association variants[98].

Regulatory elements such as enhancers and silencers are typically located more than 1 kb away from their target genes, and regulate transcription through long-range interactions, mediated by the formation of chromatin loops[100]. Once a regulatory element and its target gene(s) have been identified, chromatin conformation capture (3C) analysis can reveal the presence of potentially allele-specific chromatin interactions between the enhancer and the promoter(s) of the target gene(s)[101], which finally provides direct evidence for a regulatory variant that was initially identified by a GWAS to modulate enhancer activity, and loop formation between the enhancer and the promoter region of its target gene, and

subsequently regulates expression of that gene.

Genetic variation in noncoding RNAs

Although most post-GWAS focus on *cis*-regulatory variation to explain phenotype-association, it is also possible that associated noncoding DNA variants affect other processes like RNA splicing or gene expression regulation mediated by noncoding RNAs (ncRNAs). These regulatory ncRNAs can be separated into two categories; small RNAs (< 200 bp) such as microRNAs (miRNAs), and long noncoding RNAs (lncRNAs; > 200 bp). Whether all ncRNAs are functional, and if so, their precise biological function remains to be determined. Numerous studies reported the involvement of ncRNAs in a broad variety of biological functions[102], for example, miRNAs are implicated to function as post-transcriptional repressors by binding to 3'UTRs of target mRNAs[103]. Recent studies indicated that genetic variation directly changing the miRNA sequence or modifying their complementary sequence on target mRNAs affects miRNA repressive functions, resulting into altered target gene expression and the associated phenotype[104,105].

With the emergence of next-generation deep sequencing many lncRNAs that were initially considered to be junk RNA transcribed from junk DNA, were identified by, for example, the ENCODE project (>9.000 lncRNAs). Moreover, numerous reports assigned different functions to individual lncRNAs, including involvement in (different levels of) transcriptional regulation, scaffolding, and organization of chromatin and nuclear structure[102]. Some lncRNAs are transcribed from intergenic regions of the human genome[106], and it was reported recently that these regions often overlap with genomic regions that are associated with disease risk[107]. Genetic variants might, for example, alter the lcnRNA sequence, or modulate expression of lncRNAs[107–109]. Alternatively, genetic variants might alter the complementary sequence of the target gene. These scenarios all affect lncRNA function and subsequent transcriptional regulation of target genes, possibly resulting into the associated phenotype. Studying noncoding RNAs, and the involvement of genetic variations herein, requires another level of sequencing technologies, as for example standard RNA-seq is not sufficient to identify rare transcripts such as lcnRNAs. Following identification of lncRNAs potentially affected by genetic variants, combination of methods, comparable with the functional approach for regulatory association variants can be applied to elucidate the underlying mechanism. Furthermore, it was recently discovered that another class of lncRNAs are transcribed directly from transcriptional enhancer elements (eRNAs), and it is suggested that these eRNAs are involved in the binding of the Mediator complex to bridge to the core promoter of the target gene, securing a stable transcription initiation process[110,111].

Genetic variation in coding regions with regulatory potential

So far, coding SNPs were only considered for their potential to alter protein function or structure; however, several studies[112–117] provide evidence that this assumption needs some reevaluation. The coding proportion of the human genome is highly redundant, in which most amino acids can be defined by two or more synonymous codons[118]. The observed ratios of synonymous codons are however highly nonrandom and the choice of codons used

is greatly biased in some proteins, suggesting additional regulatory mechanisms operating in certain protein-coding areas of the genome[118–120]. Recently, Stergachis et al[112] found that about 15% of human codons are dual-use codons ('duons') that simultaneously specify both amino acids and TF recognition sites. Furthermore, more than 17% of DNA variants that are associated with a trait or disease phenotype are located in these regions and potentially affect transcription factor binding. This is particularly interesting for those coding association variants that do not alter protein-codes, instead they might modulate regulatory activities of a duon. The precise mechanisms of these exonic regulatory elements are not yet unraveled, but one could think of triggering alternative promoters with different transcriptional start sites resulting into alternative protein products, acting as a distal enhancer element for a neighboring gene or affecting higher-order chromatin structures. Thus, when studying the functional impact of genetic associations that are located in exonic regions, one should not only focus on the protein code, but also on the regulatory potential of that region.

Biological function of the target gene

The final step in going from GWAS to function is to establish the biological role of the genes regulated and/or affected by the identified functional SNPs. Exploring the biological function of the proteins encoded by the target genes is undeniably crucial, as these proteins are directly determining the biological function that underlies the associated phenotype and will provide valuable insights in the fundamental mechanisms of human complex traits, such as pigmentation phenotypes, and diseases. Various experimental approaches using cell lines, primary tissues and in vivo models will need to be applied to fully elucidate the biological function and the involvement of the target gene in the studied phenotype.

Concluding remarks

The work I describe in this thesis provides and applies a detailed step-wise strategy to elucidate the functional biology underlying genetic association signals, using pigmentation traits as a phenotype and study model. Even though the described strategy might provide strong (experimental) evidence that a SNP is the direct cause of any given phenotypic association, it is however unlikely that absolute and definitive proof of such a causality of a noncoding, regulatory SNP will be found. This is most probably due to the fact that a complex phenotype such as pigmentation is not caused by one particular SNP, but the result of combinations of causal SNPs, (tissue-specific) regulatory elements and target genes. Additionally, SNPs can act in unexpected cell types or be involved in yet undefined mechanisms of which the effects cannot yet be anticipated. It is therefore going to be difficult to determine the exact contribution of one particular SNP. Nonetheless, the knowledge that can be obtained at present using the described research strategy as I demonstrated with this thesis, paves the way to discover and reveal more and more complex mechanisms, bringing us closer to elucidating the genetic basis of complex traits and diseases. It also allows us to understand why the SNPs studied in this thesis carry a (strong) predictive effect.

References

1. Jablonski, N. G. & Chaplin, G. The evolution of human skin coloration. *J. Hum. Evol.* **39,** 57–106 (2000).
2. Parra, E. J. Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am. J. Phys. Anthropol.* **Suppl 45,** 85–105 (2007).
3. Rees, J. L. & Harding, R. M. Understanding the evolution of human pigmentation: recent contributions from population genetics. *J. Invest. Dermatol.* **132,** 846–53 (2012).
4. Liu, F., Wen, B. & Kayser, M. Colorful DNA polymorphisms in humans. *Semin. Cell Dev. Biol.* **24,** 562–75 (2013).
5. Eiberg, H. *et al.* 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. Genet.* **123,** 177–87 (2008).
6. Sturm, R. a *et al.* A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82,** 424–31 (2008).
7. Liu, F. *et al.* Eye color and the prediction of complex phenotypes from genotypes. *Curr. Biol.* **19,** R192–3 (2009).
8. Jacobs, L. C. *et al.* Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Hum. Genet.* **132,** 147–158 (2013).
9. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6,** e1000993 (2010).
10. Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4,** e1000074 (2008).
11. Liu, F. *et al.* Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet.* **6,** e1000934 (2010).
12. Nan, H. *et al.* Genome-wide association study of tanning phenotype in a population of European ancestry. *J. Invest. Dermatol.* **129,** 2250–7 (2009).
13. Larue, L., Kumasaka, M. & Goding, C. R. Beta-catenin in the melanocyte lineage. *Pigment Cell Res.* **16,** 312–317 (2003).
14. Levy, C., Khaled, M. & Fisher, D. E. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol. Med.* **12,** 406–14 (2006).
15. Devine, J. H., Hewetson, A., Lee, V. H. & Chilton, B. S. After chromatin is SWItched-on can it be RUSHed? *Molecular and Cellular Endocrinology* **151,** 49–56 (1999).
16. Yasumoto, K., Takeda, K., Saito, H. & Watanabe, K. Microphthalmia-associated transcription factor interacts with LEF-1 , a mediator of Wnt signaling. *EMBO J.* **21,** 2703–2714 (2002).
17. Yun, K., So, J.-S., Jash, A. & Im, S.-H. Lymphoid enhancer binding factor 1 regulates transcription through gene looping. *J. Immunol.* **183,** 5129–5137 (2009).
18. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–192 (2013).
19. Rosenbloom, K. R. *et al.* ENCODE Data in the UCSC Genome Browser: Year 5 update. *Nucleic Acids Res.* **41,** (2013).
20. Strub, T. *et al.* Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30,** 2319–32 (2011).
21. Li, J. *et al.* YY1 regulates melanocyte development and function by cooperating with MITF. *PLoS Genet.* **8,** e1002688 (2012).
22. Donnelly, M. P. *et al.* A global view of the OCA2-HERC2 region and pigmentation. *Hum. Genet.* **131,** 683–696 (2012).
23. Duffy, D. L. *et al.* A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am. J. Hum. Genet.* **80,** 241–252 (2007).
24. Kayser, M. *et al.* Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am. J. Hum. Genet.* **82,** 411–23 (2008).
25. Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39,** 1443–52 (2007).
26. Beleza, S. *et al.* Genetic Architecture of Skin and Eye Color in an African-European Admixed Population. *PLoS Genet.* **9,** (2013).

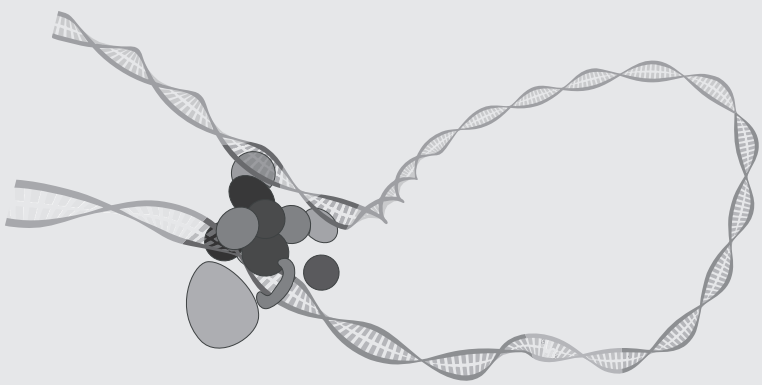
27. Cook, A. L. *et al.* Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. *J. Invest. Dermatol.* **129,** 392–405 (2009).
28. Prota, G., Hu, D. N., Vincensi, M. R., McCormick, S. a & Napolitano, a. Characterization of melanins in human irides and cultured uveal melanocytes from eyes of different colors. *Exp. Eye Res.* **67,** 293–9 (1998).
29. Ito, S. & Wakamatsu, K. Chemistry of mixed melanogenesis - Pivotal roles of dopaquinone. *Photochemistry and Photobiology* **84,** 582–592 (2008).
30. Wielgus, A. R. & Sarna, T. Melanin in human irides of different color and age of donors. *Pigment Cell Res.* **18,** 454–464 (2005).
31. Vanhoutteghem, A. & Djian, P. The human basonuclin 2 gene has the potential to generate nearly 90,000 mRNA isoforms encoding over 2000 different proteins. *Genomics* **89,** 44–58 (2007).
32. Romano, R. A., Li, H., Tummala, R., Maul, R. & Sinha, S. Identification of Basonuclin2, a DNA-binding zinc-finger protein expressed in germ tissues and skin keratinocytes. *Genomics* **83,** 821–833 (2004).
33. Vanhoutteghem, A. & Djian, P. Basonuclin 2: an extremely conserved homolog of the zinc finger protein basonuclin. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3468–73 (2004).
34. Hervé, F., Vanhoutteghem, A. & Djian, P. [Basonuclins and DISCO proteins: regulators of development in vertebrates and insects]. *Med. Sci. (Paris).* **28,** 55–61 (2012).
35. Vanhoutteghem, A. & Djian, P. Basonuclins 1 and 2, whose genes share a common origin, are proteins with widely different properties and functions. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 12423–12428 (2006).
36. Song, H. *et al.* A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat. Genet.* **41,** 996–1000 (2009).
37. Wentzensen, N. *et al.* Genetic variation on 9p22 is associated with abnormal ovarian ultrasound results in the prostate, lung, colorectal, and ovarian cancer screening trial. *PLoS One* **6,** (2011).
38. Praetorius, C. *et al.* A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155,** 1022–33 (2013).
39. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
40. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
41. Lee, T. I. & Young, R. a. Transcriptional regulation and its misregulation in disease. *Cell* **152,** 1237–51 (2013).
42. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. & Zhao, K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Research* **22,** 490–503 (2012).
43. Mercer, T. R. *et al.* Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature Biotechnology* **30,** 99–104 (2011).
44. Hirobe, T. How are proliferation and differentiation of melanocytes regulated? *Pigment Cell Melanoma Res.* **24,** 462–78 (2011).
45. Rosemblat, S. *et al.* Identification of a melanosomal membrane protein encoded by the pink-eyed dilution (type II oculocutaneous albinism) gene. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12071–75 (1994).
46. Manga, P., Boissy, R. E., Pifko-Hirst, S., Zhou, B. K. & Orlow, S. J. Mislocalization of melanosomal proteins in melanocytes from mice with oculocutaneous albinism type 2. *Exp. Eye Res.* **72,** 695–710 (2001).
47. Potterf, S. B. *et al.* Normal tyrosine transport and abnormal tyrosinase routing in pink-eyed dilution melanocytes. *Exp. Cell Res.* **244,** 319–26 (1998).
48. Patterson, L. B. & Parichy, D. M. Interactions with iridophores and the tissue environment required for patterning melanophores and xanthophores during zebrafish adult pigment stripe formation. *PLoS Genet.* **9,** e1003561 (2013).
49. Furumura, M., Sakai, C., Abdel-Malek, Z., Barsh, G. S. & Hearing, V. J. The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. *Pigment Cell Res.* **9,** 191–203 (1996).
50. Suzuki, I. *et al.* Agouti signaling protein inhibits melanogenesis and the response of human melanocytes to alpha-melanotropin. *J. Invest. Dermatol.* **108,** 838–842 (1997).
51. Haltaufderhyde, K. D. & Oancea, E. Genome-wide transcriptome analysis of human epidermal

melanocytes. *Genomics* **104,** 482–9 (2014).
52. Yamaguchi, Y. & Hearing, V. J. Physiological factors that regulate skin pigmentation. *Biofactors* **35,** 193–9 (2010).
53. Nishimura, E. K. Melanocyte stem cells: a melanocyte reservoir in hair follicles for hair and skin pigmentation. *Pigment Cell Melanoma Res.* **24,** 401–10 (2011).
54. Del Marmol, V. *et al.* Glutathione Depletion Increases Tyrosinase Activity in Human Melanoma Cells. *J. Invest. Dermatol.* **101,** 871–874 (1993).
55. Ito, S. The IFPCS presidential lecture: a chemist's view of melanogenesis. *Pigment Cell Res.* **16,** 230–6 (2003).
56. Jara, J. R., Aroca, P., Solano, F., Martinez, J. H. & Lozano, J. A. The role of sulfhydryl compounds in mammalian melanogenesis: the effect of cysteine and glutathione upon tyrosinase and the intermediates of the pathway. *Biochim. Biophys. Acta* **967,** 296–303 (1988).
57. Panzella, L. *et al.* Red human hair pheomelanin is a potent pro-oxidant mediating UV-independent contributory mechanisms of melanomagenesis. *Pigment Cell Melanoma Res.* **27,** 244–252 (2014).
58. Staleva, L., Manga, P. & Orlow, S. Pink-eyed dilution protein modulates arsenic sensitivity and intracellular glutathione metabolism. *Mol. Biol. Cell* **13,** 4206–4220 (2002).
59. Bultman, S. J., Michaud, E. J. & Woychik, R. P. Molecular characterization of the mouse agouti locus. *Cell* **71,** 1195–1204 (1992).
60. Duhl, D. M. *et al.* Pleiotropic effects of the mouse lethal yellow (Ay) mutation explained by deletion of a maternally expressed gene and the simultaneous production of agouti fusion RNAs. *Development* **120,** 1695–1708 (1994).
61. Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.* **5,** 827–35 (2004).
62. Rajesh, K., Iyer, A., Suragani, R. N. V. S. & Ramaiah, K. V. A. Intersubunit and interprotein interactions of alpha- and beta-subunits of human eIF2: Effect of phosphorylation. *Biochem. Biophys. Res. Commun.* **374,** 336–40 (2008).
63. Heaney, J. D., Michelson, M. V, Youngren, K. K., Lam, M.-Y. J. & Nadeau, J. H. Deletion of eIF2beta suppresses testicular cancer incidence and causes recessive lethality in agouti-yellow mice. *Hum. Mol. Genet.* **18,** 1395–404 (2009).
64. Law, M. H., Macgregor, S. & Hayward, N. K. Melanoma genetics: recent findings take us beyond well-traveled pathways. *J. Invest. Dermatol.* **132,** 1763–74 (2012).
65. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42,** D1001–6 (2014).
66. Zhang, X., Bailey, S. D. & Lupien, M. Laying a solid foundation for Manhattan--'setting the functional basis for the post-GWAS era'. *Trends Genet.* **30,** 140–9 (2014).
67. Raychaudhuri, S. Mapping rare and common causal alleles for complex human diseases. *Cell* **147,** 57–69 (2011).
68. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22,** 1748–1759 (2012).
69. The International HapMap Project. *Nature* **426,** 789–96 (2003).
70. The 1000 Genomes Project Consortium. An integrated map of genetic variation. *Nature* **135,** 0–9 (2012).
71. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).
72. Johnson, A. D. *et al.* SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24,** 2938–2939 (2008).
73. Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9742–6 (2010).
74. Harismendy, O. *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* **470,** 264–8 (2011).
75. Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21,** 3255–63 (2012).
76. Tuupanen, S. *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.* **41,** 885–90 (2009).
77. Van den Boogaard, M. *et al.* Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J. Clin. Invest.* **122,** 2519–30 (2012).

78. Wasserman, N. F., Aneas, I. & Nobrega, M. A. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res.* **20,** 1191–7 (2010).
79. Wright, J. B., Brown, S. J. & Cole, M. D. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol. Cell. Biol.* **30,** 1411–20 (2010).
80. Praetorius, C. *et al.* A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155,** 1022–33 (2013).
81. Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A. & Kingsley, D. M. A molecular basis for classic blond hair color in Europeans. *Nat. Genet.* **46,** 748–52 (2014).
82. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15,** 272–86 (2014).
83. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28,** 1045–8 (2010).
84. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507,** 455–61 (2014).
85. Edwards, S. L., Beesley, J., French, J. D. & Dunning, M. Beyond GWASs: Illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93,** 779–797 (2013).
86. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
87. Mathelier, A. *et al.* JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42,** (2014).
88. Matys, V. TRANSFAC(R): transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31,** 374–378 (2003).
89. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501,** 506–11 (2013).
90. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human tissues: The muTHER study. *PLoS Genet.* **7,** (2011).
91. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325,** 1246–1250 (2009).
92. Fu, J. *et al.* Unraveling the Regulatory Mechanisms Underlying Tissue-Dependent Genetic Variation of Gene Expression. *PLoS Genetics* **8,** e1002431 (2012).
93. Fehrmann, R. S. N. *et al.* Trans-eqtls reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the hla. *PLoS Genet.* **7,** (2011).
94. Garnier, S. *et al.* Genome-Wide Haplotype Analysis of Cis Expression Quantitative Trait Loci in Monocytes. *PLoS Genet.* **9,** (2013).
95. Kumar, V. *et al.* Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression. *PLoS Genet.* **9,** (2013).
96. Lalonde, E. *et al.* RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res.* **21,** 545–554 (2011).
97. Pai, A. A. *et al.* The Contribution of RNA Decay Quantitative Trait Loci to Inter-Individual Variation in Steady-State Gene Expression Levels. *PLoS Genet.* **8,** (2012).
98. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–23 (2013).
99. Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.* **7,** 171–92 (2012).
100. Bartkuhn, M. & Renkawitz, R. Long range chromatin interactions involved in gene regulation. *Biochim. Biophys. Acta* **1783,** 2161–6 (2008).
101. De Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **58,** 189–91 (2012).
102. Cech, T. R. & Steitz, J. a. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157,** 77–94 (2014).
103. Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136,** 215–233 (2009).
104. Brest, P. *et al.* A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.* **43,** 242–245 (2011).
105. Thomas, R. *et al.* HLA-C cell surface expression and control of HIV/AIDS correlate with a variant

upstream of HLA-C. *Nat. Genet.* **41,** 1290–1294 (2009).
106. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482,** 339–346 (2012).
107. Cunnington, M. S., Koref, M. S., Mayosi, B. M., Burn, J. & Keavney, B. Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet.* **6,** (2010).
108. Broadbent, H. M. *et al.* Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum. Mol. Genet.* **17,** 806–814 (2008).
109. Jendrzejewski, J. *et al.* The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8646–8651 (2012).
110. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
111. Ørom, U. A. & Shiekhattar, R. Long non-coding RNAs and enhancers. *Curr. Opin. Genet. Dev.* **21,** 194–198 (2011).
112. Stergachis, A. B. *et al.* Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342,** 1367–72 (2013).
113. Hyder, S. M., Nawaz, Z., Chiappetta, C., Yokoyama, K. & Stancel, G. M. The protooncogene c-jun contains an unusual estrogen-inducible enhancer within the coding sequence. *J. Biol. Chem.* **270,** 8506–13 (1995).
114. Lang, G., Gombert, W. M. & Gould, H. J. A transcriptional regulatory element in the coding sequence of the human Bcl-2 gene. *Immunology* **114,** 25–36 (2005).
115. Ritter, D. I., Dong, Z., Guo, S. & Chuang, J. H. Transcriptional enhancers in protein-coding exons of vertebrate developmental genes. *PLoS One* **7,** e35202 (2012).
116. Khan, A. H., Lin, A. & Smith, D. J. Discovery and characterization of human exonic transcriptional regulatory elements. *PLoS One* **7,** e46098 (2012).
117. Birnbaum, R. Y. *et al.* Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* **22,** 1059–68 (2012).
118. Grantham, R., Gautier, C. & Gouy, M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* **8,** 1893–1912 (1980).
109. Itzkovitz, S., Hodis, E. & Segal, E. Overlapping codes within protein-coding sequences. *Genome Res.* **20,** 1582–1589 (2010).
120. Warnecke, T., Weber, C. C. & Hurst, L. D. Why there is more to protein evolution than protein function: splicing, nucleosomes and dual-coding sequence. *Biochem. Soc. Trans.* **37,** 756–761 (2009).

Summary / Samenvatting



Summary

Pigmentation is one of the most obvious visual characteristics found in humans. Variation in human pigmentation is mainly due to differences in amount, type and distribution of melanin synthesized in melanocytes. Many genes have been shown to be involved in the pigmentation pathways, and as a result of high-throughput screening of DNA in genome-wide association studies (GWASs), many single nucleotide polymorphisms (SNPs) were shown to be associated with pigmentation phenotypes. The majority of the SNPs identified via GWASs (targeting any given normal or disease-related phenotype, including pigmentation) are located in noncoding regions of the genome, suggesting a regulatory role. In this thesis, the functional biology behind SNPs strongly associated with human pigmentation phenotypes is studied. In **Chapter 1**, I start with introducing human pigmentation, its evolutionary history and biochemistry, and the mechanisms of melanogenesis. Followed by a background summary on the molecular genetics of human pigmentation, which includes a review of the essentials of transcriptional regulation and the key players involved herein.

In **Chapter 2** I present the elucidation of the regulatory function of rs12913832. This SNP is located in intron 86 of the *HERC2* gene and it was shown to be highly associated with human pigmentation phenotypes. Together with my collaborators, I explored the biological mode-of-action of this strong genetic association experimentally and concluded that rs12913832 is located in an enhancer element that regulates expression of the well-known neighboring pigmentation gene *OCA2*. The activity of this enhancer depends on the allelic status of rs12913832; when the rs12913832 T-allele is present, the transcription factors LEF1, MITF and HLTF are recruited by the enhancer, and a long-range chromatin loop is present between the enhancer and the *OCA2* promoter, resulting in elevated transcription of *OCA2* and consequently the dark pigmentation phenotype. In contrast, when the rs12913832 C-allele is present, transcription factor binding and loop formation are reduced, resulting into decreased *OCA2* expression and the light pigmentation phenotype.

The genetic association signals identified by GWASs, are so-called lead SNPs that tag for haplotypes on which the directly functional or causal SNPs reside. This implies that SNPs prioritized by GWASs might not be the actual functional SNPs, which should be revealed before being able to elucidate the functional biology underlying the association signal. In **Chapter 3** I addressed the biology behind the SNP rs10756819 located in the first intron of the *BNC2* gene, and together with my collaborators I found that instead of this skin pigmentation-associated SNP, a nearby intergenic SNP rs12350739 that is in high linkage disequilibrium with rs10756819, is most likely the causal SNP for the pigmentation-association signal. The region around rs12350739 is highly conserved and functions as an enhancer element regulating expression of *BNC2* in human skin melanocytes in an allele-specific manner. Additionally, I describe the identification of an alternative promoter initiating *BNC2* transcription at exon 3, and a tissue-specific enhancer element located 6 kb upstream of this alternative promoter. I suggest that the genotype-dependent enhancer at rs12350739 and the tissue-specific intronic enhancer cooperatively boost *BNC2* transcription by both acting on the alternative promoter, resulting into a differentially expressed isoform that has a specific function in pigmentation pathways.

Regulatory elements are typically located at large distances from their target

genes, however this generally does not affect the activity of these elements, as they are able to regulate transcription through long-range interactions, mediated by the formation of chromatin loops. Focusing on chromatin structure, I characterized, together with my collaborators, the allele-specific regulatory function of an intronic enhancer that controls expression of the *IRF4* gene in **Chapter 4**. We found that irrespective of the trans-acting environment, the activity and the chromatin features of this intronic enhancer strongly depend on the allelic status of the SNP rs12203592 that it contains. I also demonstrated that this rs12203592 enhancer physically interacts with the *IRF4* promoter through an allele-dependent loop, and I suggest that subsequent allele-specific activation of *IRF4* expression is stabilized by another allele-specific loop from the rs12203592 enhancer to an additional regulatory element in intron 7 of *IRF4*.

Although it might be concluded from the studies described in **Chapter 2, 3 and 4** that the functional characterization of genetic association signals can be relatively straightforward, the study in **Chapter 5** however exemplifies that this can be rather complex. In this chapter I described the identification of five genomic regions harboring SNPs that are associated at a genome-wide significant level with skin color phenotypes. The molecular biology behind the identified SNPs in the 20q22.11 region, representing the least understood of the 5 identified genomic regions, was experimentally investigated in detail. Together with my collaborators, I analyzed transcription patterns of all genes spanning the association signals at 20q22.11 in relevant tissue and cell lines; and correlated this with their respective pigmentation phenotypes and with the genotypes of the identified pigmentation-associated SNPs. Profiling the chromatin at 20q22.11 for features of regulatory elements revealed that the region is transcriptionally active in epidermal cells; however, none of the regions identified with regulatory potential coincided with the physical locations of the identified pigmentation-associated DNA variants. Combining the obtained functional data prioritized two genes that are likely to be involved in pigmentation pathways; *EIF2S2* and *GSS*. Additionally, evidence is provided that *ASIP*, a gene well-known to be involved in pigmentation pathways, acts on the regulation of skin pigmentation from the skin dermis, rather than from the skin epidermis.

In summary, I describe the transcriptional regulation of several human pigmentation genes mediated by enhancer elements that harbor pigmentation-associated (LD-)SNPs. Activity of the enhancers, and subsequently the expression level of the genes that are controlled by these enhancers, are modulated depending on the allelic status of these SNPs. Furthermore, GWASs identified a novel SNP associated with skin color differences at region 20q22.11, and, in combination with functional follow-up, this lead to the identification of two novel genes that are likely to be involved in pigmentation. Finally, in **chapter 6**, I discuss the results of **chapters 2-5** in a more general context and describe how the highly valuable information obtained with GWASs can be functionally interpreted to unravel piece-by-piece the complex pathways of transcriptional regulation, reflecting important steps to be taken from statistical genetic association towards understanding the biological function. This is not only useful in deciphering the functional biology of SNPs associated with common phenotypes like human pigmentation, but also in studying the biology behind disease-related SNPs. It also provides a functional explanation on why the SNPs studied here are suitable to predict human pigmentation traits from DNA.

Samenvatting

De pigmentatie van de mens is één van onze meest opvallende uiterlijke kenmerken. De diversiteit in humane pigmentatie wordt voornamelijk bepaald door de variatie in hoeveelheid, het type en de distributie van melanine dat in melanocyten wordt geproduceerd. Voor diverse genen is bewezen dat ze betrokken zijn bij pigmentatie pathways. Door middel van het high-throughput screenen van DNA in genoom-brede associatie studies (GWASs), is aangetoond dat ook veel single nucleotide polymorfismes (SNPs) gerelateerd zijn met pigmentatie-fenotypes. Het overgrote deel van de SNPs die geïdentificeerd zijn via GWASs (gericht op elk willekeurig normaal of ziekte-gerelateerd fenotype, inclusief pigmentatie) bevinden zich in non-coderende regio's van het genoom, en dit suggereert een regulerende rol voor deze SNPs. In dit proefschrift is de functionele biologie achter de DNA varianten die sterk geassocieerd zijn met humane huidskleur, bestudeerd. In **Hoofdstuk 1** begin ik met een introductie van humane pigmentatie; de evolutionaire geschiedenis, de biochemie en de mechanismes van melanogenesis. Vervolgens geef ik een samenvatting van de moleculaire genetica van humane pigmentatie, waarin ik ook de essenties van transcriptionele regulatie en de belangrijkste spelers die hierbij betrokken zijn, bespreek.

In **Hoofdstuk 2** presenteer ik de regulerende rol van rs12913832. Deze SNP is gelegen in intron 86 van het *HERC2* gen, en het was aangetoond dat het sterk geassocieerd is met humane pigmentatie fenotypes. Door middel van verschillende experimenten, heb ik samen met mijn onderzoekspartners de biologische functie van deze sterke, genetische associatie uitgezocht, waarbij ik tot de conclusie kom dat rs12913832 gelegen is in een enhancer ('versterker') element dat de expressie van het naastgelegen welbekende pigmentatie gen *OCA2* reguleert. De activiteit van deze enhancer is afhankelijk van de allel-status van rs12913832; wanneer het rs12913832 T-allel aanwezig is worden de transcriptie factoren LEF1, MITF en HLTF aangetrokken door de enhancer en is er een lange-afstand chromatine-lus aanwezig tussen de enhancer en de *OCA2* promotor. Dit resulteert in verhoogde transcriptie van *OCA2* en zodoende het donkere pigmentatie-fenotype. Het tegenovergestelde vindt plaats wanneer het rs12913832 C-allel aanwezig is; de transcriptie factoren worden verminderd aangetrokken en de chromatine-lus formatie is afgenomen, en dit resulteert in een verminderde *OCA2* expressie en het lichte pigmentatie-fenotype.

De genetische associatie signalen, die geïdentificeerd worden in GWASs, zijn zogenaamde aanwijzende SNPs die haplotypes markeren waar de direct functionele of causale SNPs zich bevinden. Dit impliceert dat SNPs die geprioriteerd worden door GWASs niet per se de eigenlijke functionele varianten hoeven te zijn. Dit dient te worden onderzocht, voordat de functionele biologie dat ten grondslag ligt aan het associatie signaal kan worden opgehelderd. In **Hoofdstuk 3** richt ik me op de biologie achter de SNP rs10756819 gelegen in het eerste intron van *BNC2.* Samen met mijn onderzoek partners, heb ik ontdekt dat in plaats van deze pigmentatie-gerelateerde variant, een nabijgelegen intergene SNP; rs12350739 – dat in hoog 'gekoppeld disbalans' ('linkage disequilibrium'(LD)) is met rs10756819 – hoogstwaarschijnlijk de daadwerkelijk causale DNA variant is voor het pigmentatie associatie signaal. De regio rondom rs12350739 is sterk evolutionair geconserveerd en functioneert als een enhancer element dat, op een allel-specifieke manier, de expressie van *BNC2* reguleert in humane huidmelanocyten. Verder beschrijf ik de identificatie van een

alternatieve promotor welke de *BNC2* transcriptie initieert vanaf exon 3, en een weefsel-specifiek enhancer element dat 6 kb opwaarts ligt vanaf de alternatieve promotor. Ik suggereer dat de genotype-afhankelijke enhancer bij rs12350739 en de weefsel-specifieke enhancer in intron 2, coöperatief transcriptie van *BNC2* stimuleren, door beiden invloed uit te oefenen op de alternatieve promotor. Dit resulteert in een *BNC2*-isoform dat differentieel tot expressie komt met een specifieke functie in de pigmentatie biologie.

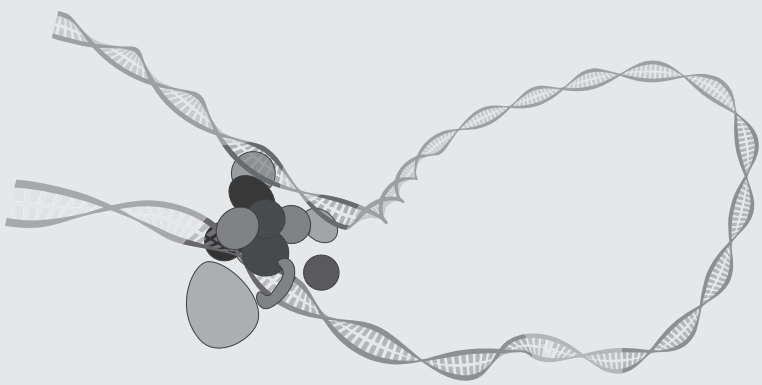
Regulerende elementen liggen vaak op grote afstand van hun doelwit genen, maar dit beïnvloedt de activiteit van deze elementen gewoonlijk niet, omdat ze in staat zijn transcriptie te reguleren door middel van lange-afstand interacties die tot stand komen door chromatine lussen. In **Hoofdstuk 4** beschrijf ik hoe een intronische enhancer de expressie van het *IRF4* gen controleert, waarbij de activiteit en de chromatine-eigenschappen van deze enhancer onafhankelijk zijn van de trans-agerende omgeving, maar daarentegen wel sterk bepaald worden door de allel-status van de SNP rs12203592, welke zich in de enhancer bevindt. Verder beschrijf ik dat deze r12203592-enhancer fysiek contact maakt met de promotor van *IRF4* door middel van een allel-afhankelijke chromatine lus, en suggereer ik dat allel-specifieke activatie van *IRF4* expressie gestabiliseerd wordt door een andere allel-specifieke chromatine lus tussen de rs12203592-enhancer en een alternatief regulerend element in intron 7 van *IRF4*.

Alhoewel vanuit de studies beschreven in **Hoofdstuk 2, 3 en 4**, geconcludeerd kan worden dat de functionele karakterisatie van genetische associatie signalen redelijk rechttoe-rechtaan is, laat de studie beschreven in **Hoofdstuk 5** duidelijk zien dat dit behoorlijk complex kan zijn. In dit hoofdstuk beschrijf ik de identificatie van 5 genomische regio's die SNPs herbergen, en die significant geassocieerd zijn (op een genoom-breed niveau) met huidskleur fenotypes. De moleculaire biologie achter de geïdentificeerde SNPs in één van de 5 regio's dat vooralsnog het minst verklaard is, de 20q22.11 regio, is experimenteel in detail onderzocht. We hebben de expressie patronen geanalyseerd van alle genen waarin de associatie signalen in de 20q22.11 regio gelegen zijn in relevante weefsels en cel lijnen, en we hebben deze gen expressie patronen gecorreleerd met de respectievelijke pigmentatie fenotypes en met de genotypes van de geïdentificeerde DNA varianten. Het chromatine profiel van de 20q22.11 regio met kenmerken van regulerende elementen onthulde dat de regio transcriptioneel actief is in zowel epidermale melanocyten als epidermale keratinocyten. Desalniettemin viel geen enkele regio dat geïdentificeerd was als potentieel regulatoir samen met de fysieke locatie van de geïdentificeerde SNPs die geassocieerd zijn met huidpigmentatie. Door de verkregen data te combineren konden twee genen, namelijk *EIF2S2* en *GSS*, worden geprioriteerd als genen die waarschijnlijk betrokken zijn bij pigmentatie pathways. Daarbij laat ik ook zien dat *ASIP*, een welbekend pigmentatie gen, invloed uitoefent op de regulatie van huidpigmentatie vanuit de dermis en niet vanuit de epidermis van de huid.

Samenvattend; ik beschrijf hoe de transcriptionele regulatie van verscheidene pigmentatie genen beïnvloed wordt door enhancer elementen waarin pigmentatie-geassocieerde (LD-)SNPs gelegen zijn. Deze SNPs beïnvloeden de activiteit van deze enhancers, en dientengevolge de expressie niveaus van de genen die onder controle staan van deze enhancers, op allel-specifieke wijze. Bovendien werd door middel van GWASs een SNP in de 20q22.11 regio geïdentificeerd als geassocieerd met huidskleurverschillen,

en in combinatie met functionele vervolganalyses resulteerde dit in de identificatie van twee nieuwe genen die waarschijnlijk betrokken zijn in pigmentatie biologie. Ik besluit dit proefschrift in **Hoofdstuk 6** met een algemene discussie waarin ik de resultaten van **Hoofdstuk 2-5** in een meer algemene context plaats. Ook beschrijf ik hoe de waardevolle informatie die verkregen wordt met GWASs gebruikt kan worden om stukje bij beetje de complexe pathways van transcriptionele regulatie te ontrafelen. Tevens worden essentiële stappen besproken, die genomen moeten worden om vanuit een statistische genetische associatie, de onderliggende biologische functie te begrijpen. Dit is niet alleen bruikbaar bij het ontcijferen van de functionele biologie achter de SNPs die geassocieerd zijn met algemene fenotypes zoals pigmentatie van de mens, maar ook bij het bestuderen van de biologie achter ziekte-gerelateerde DNA varianten. Het levert ook een biologische verklaring op waarom de SNPs die in dit proefschrift bestudeerd zijn, geschikt zijn om pigmentatie eigenschappen te voorspellen.

Curriculum Vitae & PhD Portfolio



Curriculum Vitae

Personal details

Name	Mijke Visser
Date/Place of birth	22-08-1979, in Leiden
Nationality	Dutch
Email	mijkevisser@gmail.com

Education

2008 – 2014	**PhD student** Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands PhD thesis: *From GWAS to function: transcriptional regulation of pigmentation genes in humans.*
2006 – 2007	**Master of Science in Chemistry, track Biological Chemistry** Leiden University, Leiden, The Netherlands Master Diploma Thesis (6 months): *14-3-3ζ, an oncogene?*
2000 – 2006	**Bachelor of Science in Chemistry** Leiden University, Leiden, The Netherlands Bachelor Diploma Thesis (3 months): *An attempt to knockdown 14-3-3*α *in a human cell line*

Work experience

2008 – 2014	**PhD research** Dissertation title: *From GWAS to function: transcriptional regulation of pigmentation genes in humans.* Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands Promoter: Prof. dr. Manfred Kayser & co-promoter: Dr. Robert-Jan Palstra
2006 – 2007 (9 months)	**MSc research** Master thesis title: *14-3-3ζ, an oncogene?* Department of Molecular Biology, Leiden University, Leiden, The Netherlands
2006 (3 months)	**BSc research** Bachelor thesis title: *An attempt to knockdown 14-3-3*α *in a human cell line* Department of Molecular Biology, Leiden University, Leiden, The Netherlands

Publications

1. Fan Liu, **Mijke Visser**, David L. Duffy, Pirro G. Hysi, Leonie C. Jacobs, Oscar Lao, Kaiyin Zhong, Susan Walsh, Lakshmi Chaitanya, Andreas Wollstein, Gu Zhu, Grant W. Montgomery, Anjali K. Henders, Massimo Mangino, Daniel Glass, Veronique Bataille, Richard A. Sturm, Fernando Rivadeneira, Albert Hofman, Wilfred F.J. van IJcken, André G. Uitterlinden, Robert-Jan T.S. Palstra, Timothy D. Spector, Nicholas G. Martin, Tamar E.C. Nijsten, and Manfred Kayser, for the International Visible Trait Genetics (VisiGen) Consortium. Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Submitted*
2. **M. Visser**, R-J. Palstra, M. Kayser (2015). Allele specific transcriptional regulation of *IRF4* in melanocytes is mediated by chromatin-looping of the intronic rs12203592 enhancer to the *IRF4* promoter. *Human Molecular Genetics*, in press.
3. **M. Visser**, R-J. Palstra, M. Kayser (2014). Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby *BNC2* pigmentation gene. *Human Molecular Genetics*, 23(21):5750-5762.
4. **M. Visser**, M. Kayser, F. Grosveld, R-J. Palstra (2014). Genetic variation in regulatory DNA elements: the case of OCA2 transcriptional regulation. *Pigment Cell Melanoma Research*, 27(2):169-77.
5. P.A. Lindenbergh, M. de Pagter, G. Ramdayal, **M. Visser**, D. Zubakov, M. Kayser, T. Sijen (2012). A multiplex (m)RNA profiling system for the forensic identification of body fluids and contact traces. *Forensic Science International: Genetics*, 6:565–577.
6. **M. Visser**, M. Kayser, R-J. Palstra (2012). Functional genetics of common phenotypes: *HERC2* rs12913832 modulates human pigmentation by attenuating a long-range enhancer of *OCA2* expression. *Genome Research*, 22(3):446-55.
   Featured in *Nature Reviews Genetics* 13, 148-149, 2012, Hannah Stower 'Gene expression: Looping together the function of SNPs', doi:10.1038/nrg3180
7. **M. Visser**, D. Zubakov, K.N. Ballantyne, M. Kayser (2011). mRNA-based skin identification for forensic applications. *International Journal of Legal Medicine*, 125:253–263.
8. C. Backendorf, A.E. Visser, A.G. de Boer, R. Zimmerman, **M. Visser**, P. Voskamp, Y.H. Zhang, M. Noteborn (2008). Apoptin: Therapeutic potential of an early sensor of carcinogenic transformation. *Annual Review of Pharmacology and Toxicology*, 48:143-169.
9. M. Niemantsverdriet, K. Wagner, **M. Visser**, C. Backendorf (2008). Cellular functions of 14-3-3ζ in apoptosis and cell adhesion emphasize its oncogenic character. *Oncogene*, 27:1315-1319.

PhD portfolio

Name: Mijke Visser
Department: Forensic Molecular Biology
Research School: Medical Genetics Centre (MGC)

PhD period: 2008 – 2014
Promoter: Prof. Dr. M. Kayser
Co-promoter: Dr. R-J. Palstra

PhD Training & Teaching

General academic and research skills

2009	Annual course Molecular and Cell Biology
2011	English biomedical writing and communication
2009	Safely working in the laboratory

In-depth courses

2008	Applied bioinformatics 'finding your way in biological information'
2008	Whole transcript expression microarray workshop (Leuven)
2008	Real-Time PCR training workshop
2008	Basic data analysis on gene expression arrays (BAGE)
2009	Analysis of microarray gene expression data
2009	MathWorks training on 'Matlab Fundamentals' and 'Statistical Methods'
2010	SNP course VII
2012	Epigenetic Regulation in Health and Disease
2014	Partek Training Course
2014	Photoshop and Illustrator CS6 Workshop / Indesign CS6 Workshop

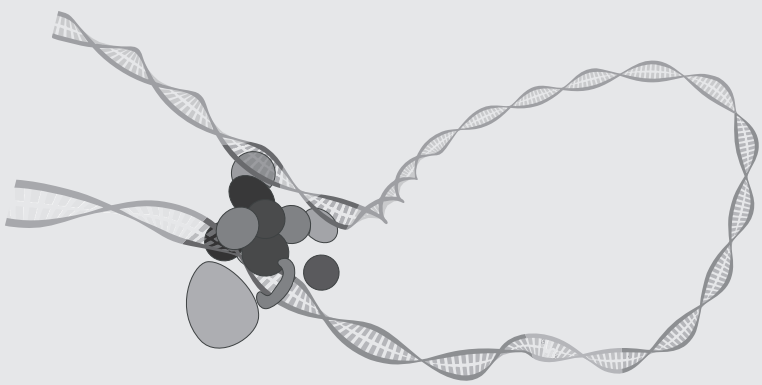
(Inter)national conferences, seminars and workshops

2008-2011	The Annual Medical Genetics Centre (MGC) symposia *(2011: oral presentation)*
2009	16th PhD workshop Bruges
2009	7th Dutch chromatin meeting in Utrecht
2009-2011	The Forensic Genomics Consortium Netherlands (FGCN) yearly seminars
2010	8th Dutch chromatin meeting in Leiden *(poster presentation)*
2011	18th PhD workshop Maastricht *(oral presentation)*
2011	The 24th International Society for Forensic Genetics (ISFG) Congress 2011 in Vienna, Austria *(poster presentation)*
2011	9th Dutch chromatin meeting in Groningen *(poster presentation)*
2012	The 6th European Academy of Forensic Science (EAFS) Conference in The Hague
2012	10th Dutch chromatin meeting in Amsterdam
2013	11th Dutch Chromatin Meeting in Rotterdam
2014	12th Dutch chromatin meeting in Utrecht

Teaching activities

2012-2013	Supervision of HBO student

Dankwoord / Acknowledgements



Tot slot het laatste hoofdstuk van mijn proefschrift. Toen ik begon aan mijn promotie traject vroeg ik me regelmatig af hoe ik het allemaal voor elkaar zou krijgen; de projecten, de artikelen, het proefschrift… Maar nu kan ik vol trots zeggen, hier ligt ie dan! Waaraan ik gelijk wil toevoegen dat dat zonder de hulp, samenwerkingen en bijdragen van zovelen in de afgelopen jaren niet mogelijk was geweest, en ik wil iedereen daar dan ook heel erg voor bedanken.

Allereerst uiteraard mijn promotor; Prof Dr. Manfred Kayser. Dear Manfred, thank you for giving me the opportunity to be a PhD student within the FMB department. I have learned so much from being here, and I am really grateful that you gave me the chance to develop myself as a scientist and to choose different research directions than originally planned. I would also like to thank you for all the time and effort that you took to discuss my work, correct my papers and the chapters of this thesis, and to discuss my career plans.

En mijn copromotor; Dr R-J. Palstra! Robert-Jan, of zoals altijd, RJ, wat begon als een (klein) samenwerkingsproject is uiteindelijk uitgemond in niet één, maar in meerdere, prachtige publicaties, èn een volledig proefschrift, met jou als copromotor. Ik had het niet beter kunnen treffen. Dank voor je support en alles wat je me hebt geleerd, voor de vele werkoverleggen, de tips in het lab, de correcties op mijn schrijfsels, maar ook voor de gezelligheid; de koffies, de lunches. Ik vind het ontzettend jammer dat onze samenwerking hier lijkt te eindigen, hopelijk lukt het ons om op een of andere manier toch nog verder te gaan, want wat mij betreft zijn we nog lang niet klaar!

De leden van mijn kleine promotiecommissie, ofwel de leescommissie; Dr. Joost Gribnau, Prof Dr. André Uitterlinden en Dr. Raymond Poot. Beste Joost, André en Raymond, dank voor de tijd en moeite die jullie hebben genomen om mijn proefschrift te lezen. De overige leden van de grote promotiecommissie; Prof Dr. Frank Grosveld, Prof Dr. Tamar Nijsten en Prof Dr. Wouter de Laat, dank voor jullie bereidwilligheid om zitting te nemen in de grote commissie.

Onmisbaar voor dit proefschrift was de deelname en sampledonaties van vele vrijwilligers, met name de patiënten van plastische chirurgie, die bereid waren om mij te woord te staan op een vaak stressvol moment, ontzettend bedankt! Het was niet altijd makkelijk om (met name in het begin) geschikte kandidaten te vinden en om de samples te verzamelen, en ik wil daarom speciaal Dr. Han van Neck en Dr. Maerten Smit, maar verder ook alle andere chirurgen, de OK-assistenten en de verpleging bedanken voor hun praktische hulp hierbij. Verder wil ik Leonie Jacobs en Emmilia Dowlatshahi van de afdeling Dermatologie bedanken voor hun hulp om mijn verzameling samples uiteindelijk te completeren. De noodzaak voor meer samples heeft me ook buiten Rotterdam gebracht, allereerst bij onze collega's in Leiden; iedereen van het FLDO die mee heeft gedaan, bedankt! I even went to Greece; Leda, thank you so much for collecting such a nice set of samples, and for inviting me into your home. I really enjoyed the time I spend with you in Thessaloniki!

Dan nu mijn twee paranimfen. Jolien, lief zusje, mijn beslissing om deze PhD-uitdaging aan te gaan, heb ik ook aan jou te danken. Je hebt me altijd volop gesteund en gestimuleerd, of dat nu vanuit het Amsterdamse was, of vanuit Brisbane, dat heeft geen verschil gemaakt. Ik ben super blij en trots dat je vanuit Australië hierheen komt om mij als paranimf bij te staan op 18 maart!! Kate; van collega's naar vriendinnen met kids, en altijd supergezellig! Ik vind het echt heel leuk dat je naast me staat de 18e, en ik kijk uit naar onze

volgende thee'tjes, speeldates (zonder ziekenhuis bezoekjes!) en/of sushi diners!

Uiteraard wil ik al mijn (ex-) collega's van het FMB bedanken. First of all, Dmitry; thank you for supervising me the first two years of my PhD, I really learned a lot, and I am proud of the paper we published together. Even though I moved toward a different direction with my PhD, I enjoyed discussing things with you, about science, but also politics, news flashes, cars (!), kids and so many other things… Susan, thank you so much for all the (motivating) chats, either in the lab, or online, you are such a great friend, and I admire your drive and ambition! Fan, I enjoyed collaborating with you on the skin color paper, and I am sure it will find its way into a nice journal sooner or later! Oscar, thanks for being such a great colleague and teacher, not only were you available for me whenever I had questions, but basically for everyone in the lab. I wish you all the best in your new career! Mannis, het was fijn om jou als collega-PhD te hebben, heel veel succes met het afronden van je proefschrift! Last, but certainly not least; Karolina, Lakshmi and Kaiyin; thanks so much for everything, you guys are next to finish, and I am confident you will get it done (this year, Karolina & Lakshmi?!). All the other colleagues I worked with at the FMB (and I hope I did not forget anyone, but if I did, I am very sorry!); Arwin, Iris, Andreas, Silke, Kaye, Katrin, Fedde, Germaine, Mark, Ying, Miriam and all the students (especially Kimberley), thank you, it has been an absolute pleasure to work with you all! Dan nog de collega's van buiten het FMB, ik heb altijd met heel veel plezier op de verschillende verdiepingen gewerkt, met name op de zevende bij Celbiologie (thanks Ralf, Anita en Petros!), en op de zesde bij Biomics (bedankt Wilfred, Rutger, Zeliha en Christel!).

En dan mijn vrienden en familie; ontzettend bedankt voor jullie steun en interesse, en wat op z'n minst net zo belangrijk was, alle leuke dingen die we hebben ondernomen; etentjes, strandafspraken, concertjes, kroeg avondjes, vakanties, weekendjes weg en 'sherlocked' events, allemaal welkome afleiding om even los te komen van het werk, vooral het laatste half jaar.

Mam en Niek, jullie hebben me altijd ondersteund, in alles wat ik heb gedaan en doe, dat betekent heel veel voor me en ik kan jullie daar niet genoeg voor bedanken!!

En dan tot slot, mijn twee 'dreaming lights', Arjen en Evie, ik kan me geen leven zonder jullie voorstellen!!!

Thank you all! Bedankt allemaal!! :) Mijke