

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

Mijke Visser, Robert-Jan Palstra<sup>†</sup> and Manfred Kayser\*

Department of Forensic Molecular Biology, Erasmus MC University Medical Centre Rotterdam, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

Received April 23, 2014; Revised and Accepted June 6, 2014

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 rs10765819 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 rs10765819 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 (SNPs) 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 rs10765819 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

\*To whom correspondence should be addressed. Tel: +31 107038073; Fax: +31 107044575; Email: m.kayser@erasmusmc.nl

<sup>†</sup>Present address: Department of Biochemistry, Erasmus MC University Medical Centre Rotterdam, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.

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 six promoters and at least 23 (alternative) exons, therefore theoretically resulting in a large number of almost 90 000 possible isoforms, encoding for over 2000 different proteins. The main human *BNC2* isoform is predicted to encode a 1099 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 h 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, two studies investigated the functional potential of pigmentation-associated SNPs in two 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-precipitation

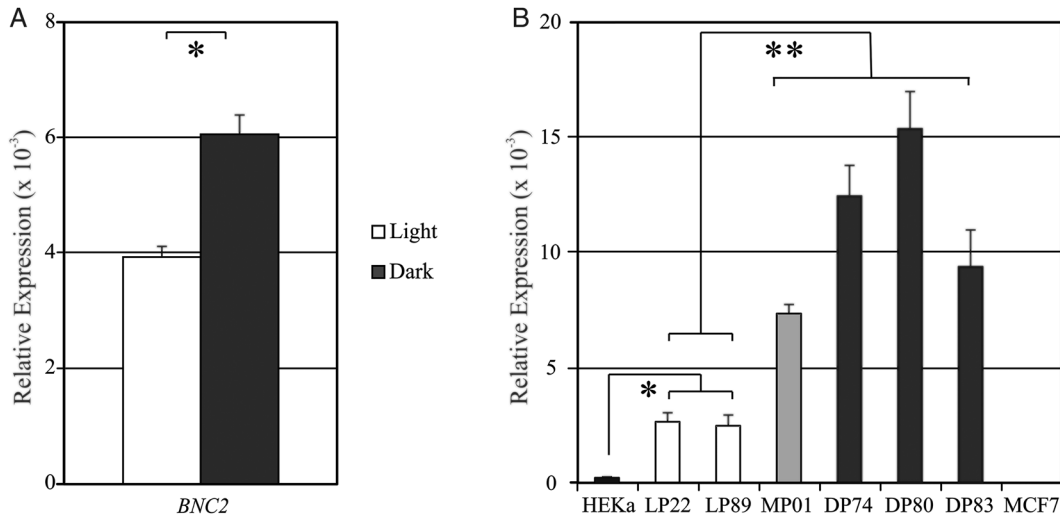
(ChIP) of H3K27Ac (an active-enhancer mark) in combination with next generation sequencing (ChIP-seq) and ENCODE data. We provide evidence that a region upstream from the canonical *BNC2* promoter acts as an enhancer regulating the expression of *BNC2*, and this transcriptional regulation is dependent 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 Supplementary Material, Fig. S1 for category separation), as well as a melanocytic cell system consisting of six commercially available primary Human Epidermal Melanocytes of neonatal origin. The melanocytes were derived from two different lightly pigmented donors (HEMn-LP22 and HEMn-LP89), from one moderately pigmented donor (HEMn-MP01) and from three darkly pigmented donors (HEMn-DP74, HEMn-DP80 and HEMn-DP83). The characterization of LP22 and DP74 was described previously (27). In addition, HRISplex analysis (30) (Supplementary Material, Tables S1 and S2) of DNA samples from the melanocytic cell lines as well as from the skin epidermal samples, and STRUCTURE analysis (31) (Supplementary Material, Fig. 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 six 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) reverse-transcriptase (RT)-PCR. We observed higher expression levels of *BNC2* in the dark skin-colored sample group as compared with 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 (~95%). Every melanocyte in the basal layer is associated with ~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



**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 six melanocyte cell lines, in keratinocytic RNA (HEK293) 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- to 70-fold lower in HEK293. 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  $\pm$  SEM; \* $P < 0.002$ ; \*\* $P < 0.00004$ . Data are represented as mean  $\pm$  SEM.

involved in freckle formation in humans by signaling from keratinocytes to melanocytes (34). In order to investigate whether the 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 (HEK293, 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, when compared with 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, the expression of *BNC2* is at least 10-fold higher in human skin melanocytes when compared with 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*.

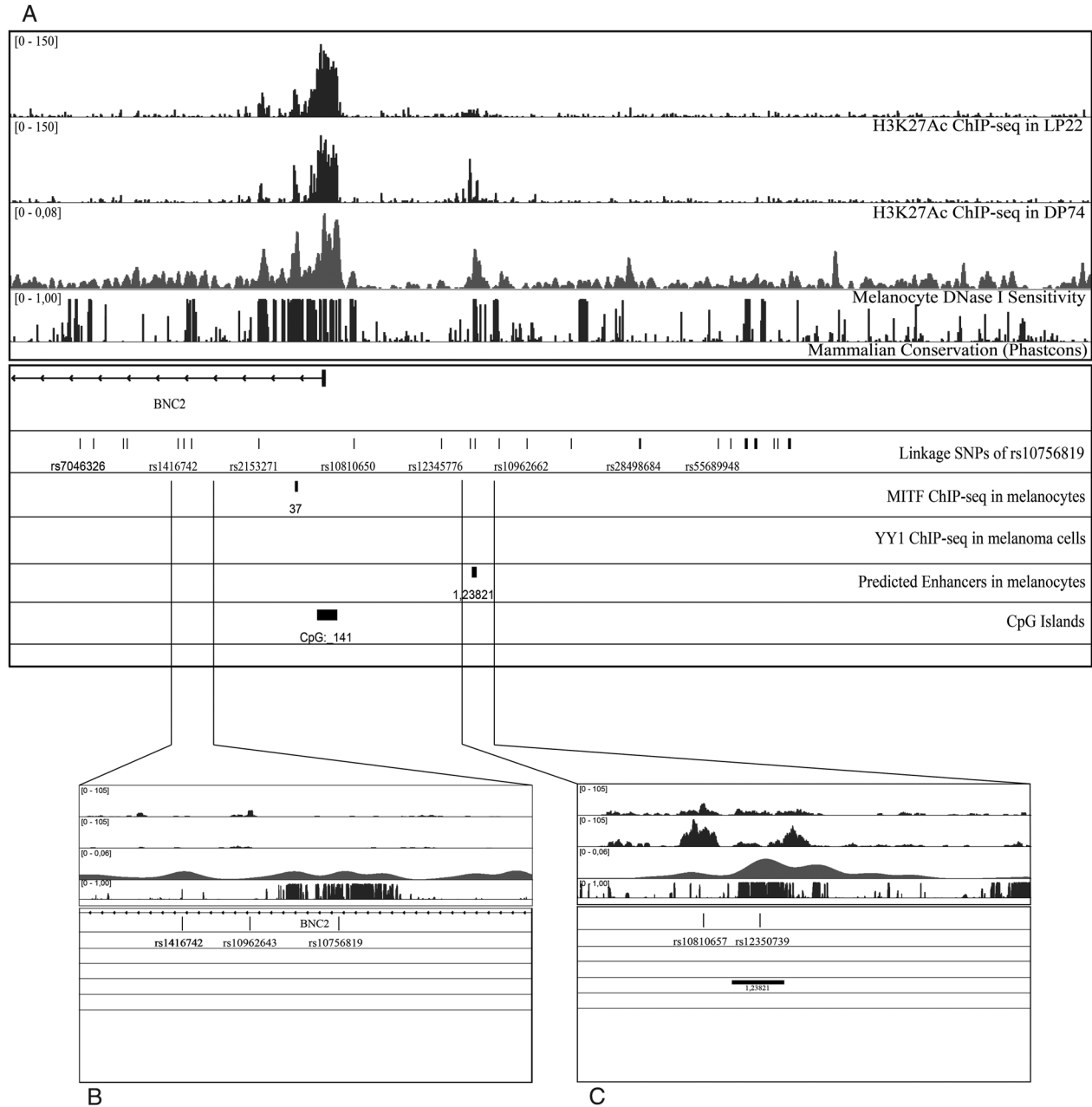
#### Allele-specific enhancer element upstream of the *BNC2* promoter

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 datasets 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 Supplementary Material, 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 (Fig. 2B).

As rs10756819 was identified as a pigmentation-associated SNP by a candidate-gene approach that only used SNPs



**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 ( $R^2 > 0.6$ ) 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 DNaseI peak is detected and an melanocyte-specific enhancer element is predicted here. Two SNPs are located in this region: rs10810657 and rs12350739.

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 (Supplementary Material, Fig. S4a, LD plot, and LD SNPs are listed in Supplementary Material,

Table S3a); these so-called LD SNPs are located within a narrow frame of 70 kb overlapping the 5' part of *BNC2* (Fig. 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

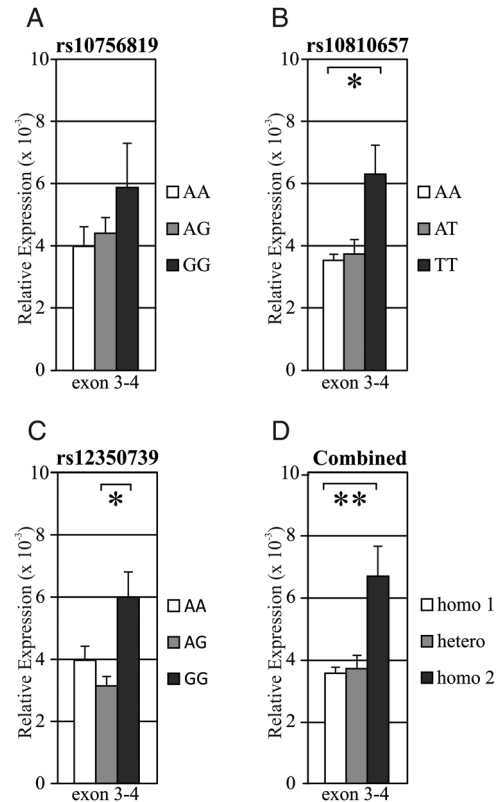
DNaseI hypersensitivity corresponding to the *BNC2* promoter region (Fig. 2A). Interestingly, two LD SNPs, rs10810657 and rs12350739, are located in a region 14 kb upstream of the first and canonical *BNC2* promoter (see Fig. 2C), which is characterized by the presence of H3K27Ac peaks that are differential between LP22 and DP74, a melanocyte DNaseI hypersensitive signal and a strong conservation signal. Furthermore, an enhancer element was predicted within this region (41) (Fig. 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 in Sub-Saharan Africans (1%), and absent in East Asians as 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 three SNPs (Supplementary Material, 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 (Fig. 3B and 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 (Fig. 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 ( $R^2 > 0.8$ , Supplementary Material, Fig. S4B and S4C and Tables S3B and C); 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 (Fig. 3D,  $P < 0.01$ ).

Taken together, these data point towards the presence of a potential regulatory element 17 kb 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

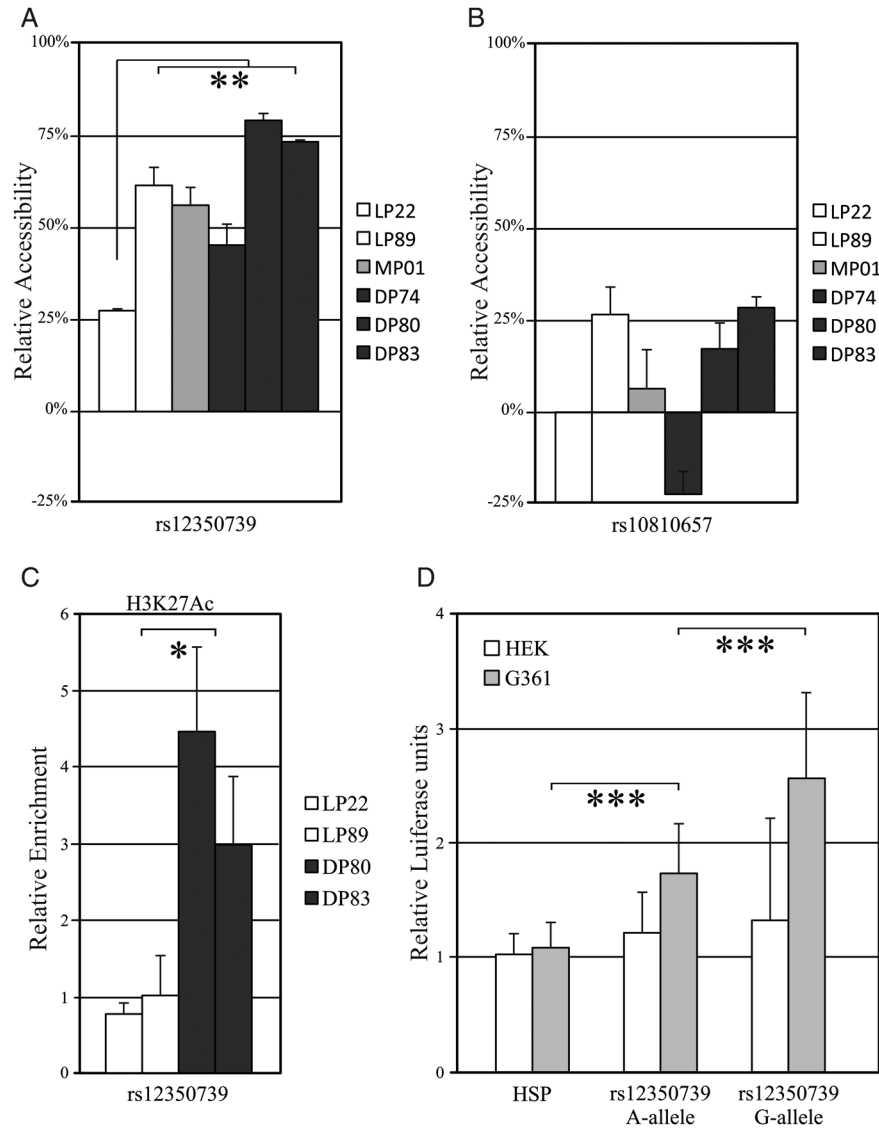
Regulatory elements are characterized by open chromatin (46). If the areas around the SNPs in this particular –17 kb region include enhancer elements, we should be able to detect accessible chromatin at these locations. We therefore used a commercially available assay that directly measures the accessibility 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,



**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  $\pm$  SEM; \* $P < 0.05$ ; \*\* $P < 0.01$ .

whereas in the cell line with the AA genotype the chromatin was fully closed (Fig. 4A). Remarkably, the area around rs10810657 was completely inaccessible in all cell lines tested (Fig. 4B). Enhancer elements characterized by a DNaseI 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 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 Fig. 2 and EpiQ results in Fig. 4B), while the position of rs12350739 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 rs12350739 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 (Fig. 2C). We confirmed the obtained H3K27Ac ChIP-seq results by ChIP-qPCR in



**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  $\pm$  SEM.

LP22 and in three additional cell lines; LP89, DP80 and DP83 (Fig. 4C). Occupancy of H3K27Ac at rs12350739 was low in both LP22 and LP89, and relatively high in DP80 and DP83 (Fig. 4C), H3K27Ac occupancy at rs12350739 corresponds 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 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,

when compared with the empty vector (Fig. 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 (Fig. 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 four regions (marked A–D in Supplementary Material, Fig. 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 four 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 (Supplementary Material, Fig. 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, ~80 kb downstream from *BNC2* (Supplementary Material, Fig. 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 an 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'' (Fig. 5A). Region B' exhibited high and differential H3K27Ac ChIP-qPCR signals among the cell lines tested; LP22 and LP89 displayed low H3K27Ac enrichment when compared with DP80 and DP83 (Fig. 5B). Notably, this corresponds with the expression pattern of *BNC2*; expression of *BNC2* is low with low H3K27Ac enrichment in LP22 and LP89, when compared with 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, 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 (Supplementary Material, Fig. S6A) and at region B' in *BNC2* (Supplementary Material, Fig. 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 (Supplementary Material, Fig. S6B). At the canonical promoter binding of RNA polymerase II is relatively low (Supplementary Material, Fig. S6A), suggesting that transcription is higher downstream of the alternative promoter. We therefore analyzed RNA-seq data we obtained from four melanocyte cell lines (LP89, DP80, DP74 and MP01). This revealed that the majority of transcripts initiated at exon 3 and not at exon 1 (Supplementary Material, Fig. S6). We confirmed this finding by RT-qPCR analysis of *BNC2*-expression levels at three different locations within the gene, using one primer set targeting mRNA upstream of the alternative promoter, and two 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, when compared with 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 Fig. 5C,  $P < 0.02$ ). Similar results were obtained when testing these three primer sets in the skin epidermal samples (Fig. 5D,  $P < 0.0004$ ).

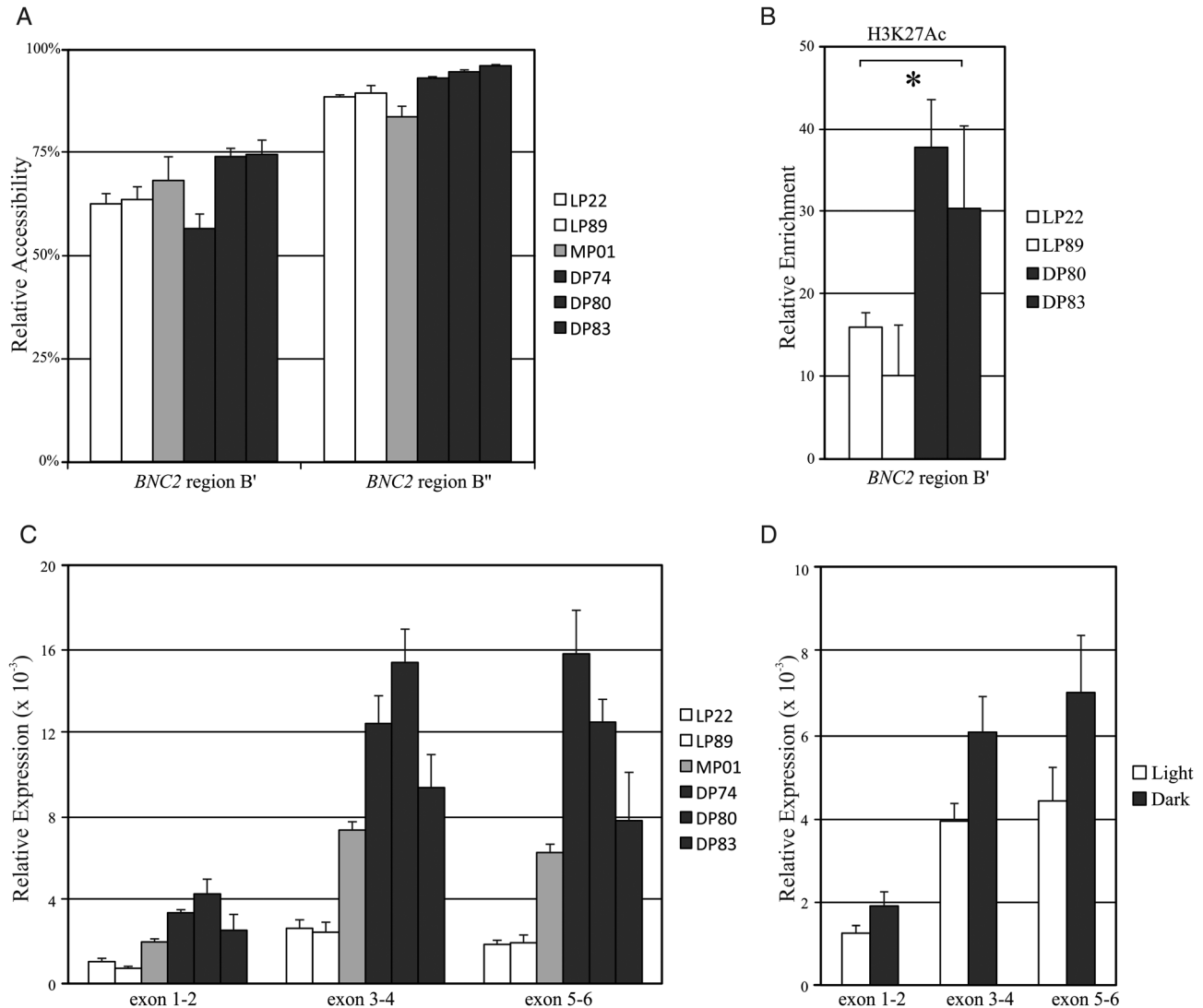
In addition, we have noted the presence of a CpG island located within the potential alternative promoter region (see Supplementary Material, Fig. S6). Unless under positive selection, methylated C nucleotides tend to be eliminated during evolution and hence CpG islands are frequently found in TATA-less 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 be 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 functions as an enhancer element for the nearby



**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 H3K27Ac-ChIP-qPCR signals in region B' are high and differential between the light- (LP22 and LP89) and dark-pigmented (DP80 and DP83) cell lines (\* $P < 0.05$ ). **(C)** RT-qPCR analysis of *BNC2* transcripts in six melanocyte cell lines at one location upstream and two 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  $\pm$  SEM.

*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 70 kb 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 with 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 when compared with those in the dark-pigmented cell lines (data not shown). EpiQ data demonstrate 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 two proposed targets *KITLG* and *CSF1* using our RNA-seq data in a subset of five 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 six promoters and many splicing isoforms, resulting in a potential of nearly 90 000 mRNA isoforms encoding >2000 different proteins (21). Our data clearly show 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 five 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 (Supplementary Material, Fig. 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°) and dark (under 28°), for a graphical display of the ITA categories of the skin epidermal sample, see Supplementary Material, 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<sup>2</sup> 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 ~1 mm wide and 7 mm long and incubating the skin pieces for 30 min 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*<sup>TM</sup> PCR Inhibitor Removal Kit, Zymo Research Corporation) following the manufacturer's instructions. Subsequent DNaseI 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% CO<sub>2</sub>. 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*<sup>TM</sup> PCR Inhibitor Removal Kit, Zymo Research Corporation) following the manufacturer's instructions. Subsequent DNaseI 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 six 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.* (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 Supplementary Material, 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 95°C, followed by 30 cycles of 15 s at 94°C and 45 s at 60°C and a final extension of 5 min at 60°C 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°C, followed by 25 cycles of 10 s at 96°C, 5 s at 50°C, and 30 s at 60°C. 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).

### ChIP-qPCR and 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 min at room temperature. To remove melanin, DNA was column purified prior to PCR (*OneStep*<sup>TM</sup> 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 Supplementary Material, 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

qPCR-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*<sup>TM</sup> 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 Supplementary Material, Table S5. Data represent at least two independent experiments. Student's two-tailed *t*-test was used to determine statistical significance.

### Transcription analysis

The RT reaction was performed using RevertAid<sup>TM</sup> 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 Supplementary Material, 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 1524 bp 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 599 bp 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 normalized to Renilla luciferase expression. Primer sequences are listed in Supplementary Material, 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 500 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).

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

### ACKNOWLEDGEMENTS

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 Emilia 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.

*Conflict of Interest statement.* None declared.

### 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., Kirsanov, 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., Pruffer, 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., Daubigny, 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.*, 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 HRisPlex 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 Steingrimsdottir, 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. Creighton, 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–D63.
  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. and 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.