PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Charanya Muralidharan

Entitled

ELUCIDATING THE MECHANISMS OR INTERACTIONS INVOLVED IN DIFFERING HAIR COLOR FOLLICLES

For the degree of Master of Science

Is approved by the final examining committee:

Dr. Susan Walsh
Chair

Dr. Lata Balakrishnan
Co-chair

Dr. Christine Picard
Co-chair

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Dr. Susan Walsh

Approved by: Dr. Stephen K. Randall 11/18/2016

Head of the Departmental Graduate Program Date

ELUCIDATING THE MECHANISMS OR INTERACTIONS INVOLVED IN DIFFERING HAIR COLOR FOLLICLES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Charanya Muralidharan

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

December 2016

Purdue University

Indianapolis, Indiana

Dedicated to my parents, my brother, both my grandmas. Each one of you follow your own dreams while inspiring, supporting and encouraging us to pursue our own.

ACKNOWLEDGEMENTS

I would not be what I am without my experiences thus far and I am grateful for each and every one that played a role. First off, I would like to begin by thanking Dr. Susan Walsh, my mentor for being an amazing professor and guide. Dr. Walsh, you truly are an inspiration and a role model for me. You are one amongst the very few people that have a great passion for science and are also passionate about spending quality time with family and loved ones. I admire your work and am so grateful and proud to be your student.

I would like to thank Dr. Christine Picard and Dr. Lata Balakrishnan for being a part of my advisory committee. You are both wonderful professors and I sincerely value your input and advise. I also would like to thank Ewelina for being such an amazing mentor. Even though, you are across the ocean and we have not met each other, you still made sure I understood what I was doing and patiently guided me. I would like to thank Dr. Lapish and Eva as well. Because of you guys, I am now much less intimidated by statistics.

I would like to extend my sincere gratitude to Krystal, Ryan, Wesli, Megan, Annie, Stephanie, Noah, Kirsten, Gina, Katherine, Mirna, Bailey and Lydia for their support and help. Krystal, you are an exceptional friend and a great support. I loved working with you and am still learning to be a confident person as you are. We make a great team and I am truly going to miss you. I am going to miss our conversations on anything and everything. You are a rock star and I hope we get to start a lab like we have always discussed! Ryan, you are like my little brother and I truly enjoyed discussing a variety of bioinformatics topics. I am going to miss your cute little post- it notes and am going to miss embarrassing you. I am sure you are going to be a great professor someday.

Wesli, you definitely are a wealth of random knowledge. I am going to miss our RNA troubleshooting days and I am going to miss our conversations on random subjects and your cheerful personality. Thank you Shari, Sue, Dr. Clark, Dean, Patrick, Asma and Pooja for their valuable guidance, help and support.

My best friends at home, Mathew, Nari, Raji, Divs- Thank you for bearing with me and being there for me even when I haven't spoken to you guys as often as I would really like to. I am thankful to my second family- Cheeku, TK, Himaja, Laks, Glad, Arun, Sumana, Ashwin, Dobby, Shravan, Vinidhra and Santosh. Cheeku, you are an incredible person and I thank you from the bottom of my heart for encouraging me to follow my passion. I cannot even try and put down in words as to how much I respect you! TK, Ashwin, Dobby and Arun, you guys are amazing! Ashwin- my stats guru, thank you for patiently helping me understand complicated stats. TK, I love our conversations on every random thing and I know I can always look up to you when I need advice. Himaja, Glad and Laks, thank you for always looking out for me and being there for me. You girls are the best! Thank you Keerthi and Anna for being such wonderful roommates. I thank everyone from ACGT, Inc. and SN for being wonderful and supportive. Thank you Usha aunty and Mouli uncle for encouraging and supporting me on my quest for knowledge!

To my family- naina, eeks (ma), buch, avva, lachu kutty, chithaps, chithi, atha, mama, Rams, Jai, Anu, maamis and kutties. I love you all so much! Thank you naina and eeks for being such wonderful parents and support system. Buch and I have such great role models to live up to and to follow our passions! I hope to make you guys proud someday and thanks to Dr. Walsh and her guidance, I know I am on the correct track. Buch, you are the best brother one could ever have. Thank you for being there for me all the time! I am thankful to avva, for, your constant questions and curiosity in genetics, helped me narrow down on the subject I want to focus on for the rest of my life. We all miss you! Lastly, I express my sincerest gratitude for all the volunteers that participated in our studies. Without your participation, I will be nowhere!

TABLE OF CONTENTS

Page

LIST OF TABLES vii
LIST OF FIGURES viii
LIST OF ABBREVIATIONS ix
ABSTRACT xi
CHAPTER 1. INTRODUCTION 1
1.1 Purpose, Goals and Objectives 1
1.2 Pigmentation Biology 4
1.3 Epistatis Investigation using select pigmentation SNPs 7
1.4 Pigmentation and Gene Expression 11
CHAPTER 2. MATERIALS & METHODS: EPISTASIS 14
2.1 Sample Collection 14
2.2 DNA Extraction, Quantitation and SNP Genotyping 14
2.3 Phenotyping 17
2.4 Statistical Analysis 19
2.4.1 Multifactor Dimensionality Reduction 23
2.4.2 Binary Logistic Regression 24
2.4.3 Model Validation 25
CHAPTER 3. MATERIALS & METHODS: HAIR FOLLICLE PIGMENT GENE EXPRESSION 27
3.1 Sample Collection 27
3.2 RNA Extraction and Quantitation 27
3.3 cDNA Synthesis 28
3.4 Quantitative PCR (qPCR) Reaction Set up 28
3.5 qPCR Data Analysis 29

Page

CHAPTER 4. RESULTS & DISCUSSION- EPISTASIS 30
4.1 Multifactor Dimensionality Reduction 30
4.2 Binary Logistic Regression 43
4.3 Predictive Modelling 44
CHAPTER 5. RESULTS & DISCUSSION: HAIR FOLLICLE PIGMENT GENE EXPRESSION 53
CHAPTER 6. CONCLUSIONS 60
REFERENCES 62
APPENDICES
Appendix A. SNaPshot Multiplex Assay Primer Information 72
Appendix B. PLINK Interaction Results 78
Appendix C. Summary of all Interactions identified via MDR and PLINK Analyses ... 105
Appendix D. Initial AUC table for all the SNPs and multiplied Diplotypes 107
Appendix E. Pearson Correlation Matrix 110
Appendix F. Sequential Model Building AUC Summary 113
Appendix G. Distribution of 3 Identified Interactions and 1 Independent SNP with Hair Color Phenotypes 114

LIST OF TABLES

Table Page

Table 2.1 Minor Allele Frequency table for 61 markers 20

Table 2.2 List of cases vs controls classifications used for MDR and BLR analysis 22

Table 4.1 MDR analysis for pairwise interactions for different trait classifications 41

Table 4.2 PLINK analysis for different trait classifications 43

Table 4.3 List of models containing SNPs/ Interactions that add on value to HIrisPlex model post cross-validation. 45

Table 4.4 AUC values (before and after cross-validation) for sequential addition of SNPs and interactions to HIrisPlex model 46

Table 4.5 Hair color prediction accuracies 48

Table 5.1 Panel of genes that were assessed using qPCR 53

LIST OF FIGURES

Figure Page

Figure 1.1 Formation of different types of melanin 5

Figure 1.2 Overview of the function of melanocytes and it's migration into keratinocytes in hair follicle pigmentation 6

Figure 2.1 Overview of SNaPshot multiplex SBE assay 16

Figure 2.2 Separation of hair colors across wavelengths (400-700nm) 18

Figure 2.3 Plot of means at 700nm for each hair color category with SD bars 19

Figure 4.1 Black Vs Non-black MDR analysis 31

Figure 4.2 Black Vs Dark Brown MDR analysis 32

Figure 4.3 Dark brown Vs Non- Dark brown MDR analysis 33

Figure 4.4 Dark brown Vs Light Brown MDR analysis 34

Figure 4.5 Light brown Vs Non- Light brown MDR analysis 35

Figure 4.6 Light brown Vs Red MDR analysis 36

Figure 4.7 Red Vs Non- red MDR analysis 37

Figure 4.8 Red Vs Blond MDR analysis 38

Figure 4.9 Blond Vs Light brown MDR analysis 39

Figure 4.10 Blond Vs Non- blond MDR analysis 40

Figure 4.11 ROC curve of hair color prediction models 47

Figure 4.12 Pigmentation pathway with proposed interactions 49

Figure 5.1 Plot of means of CT of housekeeping genes 54

Figure 5.2 Relative fold change comparison 55

LIST OF ABBREVIATIONS

SNP- Single Nucleotide Polymorphism

CODIS- Combined DNA Index System

GWAS- Genome Wide Association Studies

OCA- Oculocutaneous Albinism

HERC2- HECT and RLD domain Containing E3 Ubiquitin Protein Ligase 2

TYR- Tyrosinase

TYRP1- Tyrosinase Related Protein-1

SLC45A2- Solute Carrier family 45 member 2

IRF4- Interferon regulatory factor 4

MC1R- Melanocortin 1 receptor

MITF- Microphthalmia Transcription Factor

°C- Degree Celsius

®- Registered

™- Trademark

ng- nanogram

uL- microliter

Rpm- rotations per minute

PCR- Polymerase Chain Reaction

RFU- Relative Fluorescence Unit

SBE- Single Base Extension

nm- nanometer

LD- Linkage Disequilibrium

MDR- Multifactor dimensionality reduction

BLR- Binary Logistic Regression

MLR- Multinomial Logistic Regression

AUC- Area Under ROC Curve

ROC- Receiver Operating characteristic Curve

HKG- Housekeeping gene

TBP- TATA box binding protein

RPM- Revolutions Per Minute

ABSTRACT

Muralidharan, Charanya. M.S., Purdue University, December 2016. Elucidating the Mechanisms or Interactions Involved in Differing Hair Color Follicles. Major Professor: Susan Walsh.

Forensic DNA phenotyping is an up and coming area in forensic DNA analyses that enables the prediction of physical appearance of an individual from DNA left at a crime scene. At present, there has been substantial work performed in understanding what genes/markers are required to produce a reliable prediction of categorical eye and hair color from the DNA of an individual of interest. These pigmentation markers (variants from *HERC2, OCA2, TYR, SLC24A4, SLC45A2, IRF4* to name a few) are at the core of several prediction systems for eye and hair color such as IrisPlex, HIrisPlex, and the Snipper 2.5 suite. The contribution of these markers towards prediction in most cases however, only factors in an independent effect and do not take into account potential interactions or epistasis in the production of the final phenotypic color. Epistasis is a phenomenon that occurs when a gene's effect relies on the presence of 'modifier genes', and can display different effects (enhance/repress a particular color) in genotype combinations rather than individually.

In an effort to detect such epistatic interactions and their influence on hair color prediction models, for this current study, 872 individuals were genotyped at 61 associative and predictive pigmentation markers from several diverse population subsets. Individuals were phenotypically evaluated for eye and hair color by three separate independent assessments. Several analyses were performed using statistical approaches such as multifactor dimensionality reduction (MDR) for example, in an effort to detect if

there are any SNP- SNP epistatic interactions present that could potentially enhance eye and hair color prediction model performances. The ultimate goal of this study was to assess what SNP-SNP combinations amongst these known pigmentation genes should be included as an additional variable in future prediction models and how much they can potentially enhance overall pigmentation prediction model performance.

The second part of the project involved the analyses of several differentially expressed candidate genes between different hair color follicles of the same individual using quantitative Real Time PCR. We looked at 26 different genes identified through a concurrent non-human primate study being performed in the laboratory. The purpose of this study was to gain more insight on the level of differentially expressed mRNA between different hair color follicles within the same human individual. Data generated from this part of the project will act as a pilot study or 'proof of principle' on the mRNA expression of several pigmentation associated genes on individual beard hair of varying phenotypic colors. This analysis gives a first glimpse at expression levels that remain constant or differentiate between hairs of the same individual, therefore limiting the contribution of individual variation.

CHAPTER 1. INTRODUCTION

1.1 Purpose, Goals and Objectives

DNA, the crux of every living organism on earth is present in almost all the cells in our body. In humans, about 1% of the genome varies between individuals. It is this variability that helps us identify or predict how a person may look. The field of forensics is constantly being revolutionized, with more optimized techniques for the identification of a suspect using genetic information obtained at a crime scene. Even after decades of use, classifying Short Tandem Repeats (STRs) as the primary Forensic DNA typing analysis method still continues to be the most valuable resource in the forensic biology world [1]. However, STR typing is not of use when there is no profile match obtained in the CODIS or any forensic genetic database available to investigating authorities across the globe. To resolve this issue, few authorities suggest that DNA information for every single person needs to be added to national DNA databases. In fact, Kuwait has now made it mandatory that all of its citizens and foreign residents provide their DNA for the Kuwait national DNA database[2]. In fact, they have come up with rigorous punishments for those who don't abide by it. In the US, it is unlikely that this type of compulsory DNA data-basing will come into fruition, leaving only convicted individuals and, in some states, arrestees within the criminal database. So, in an investigation there may be cases where there is no match to this DNA database or the identification of an individual from a suspect list. In these cases, alternative DNA analyses methods may provide intelligence leads. This is where the concept of forensic DNA phenotyping comes to play. Forensic DNA phenotyping enables the prediction of a physical appearance from a sample left at a crime scene and is a relatively new and promising area of DNA analyses.

In essence this approach acts as a 'biological witness', providing information about externally visible characteristics of the suspect [3, 4]. Apart from being used in forensics, DNA phenotyping intelligence tools have a lot more practical applications and can be used for victim identification after mass disasters and putting a face to our ancestors. DNA phenotyping tools use SNP/ INDEL markers that have been previously associated with a particular externally visible trait. SNPs (Single Nucleotide Polymorphisms) are single nucleotide variations in the genome that occur either in the coding or non- coding regions of the genome. INDELs, on the other hand, denote insertions/ deletions of nucleotide(s) in the genome. These predictive markers have been developed into practical tools that could predict eye/ hair color with as little as 63 pg/uL [3].

As of now, only a handful of externally visible characteristics, such as eye color, hair color and skin color, have identified gene variants that can predict the common variation found in individuals globally from DNA. For example, variants in *MC1R* have been associated with red hair color, and variants in *SLC24A4, KITLG, and TYR*, have all been identified as containing hair color phenotype predicting variants for other categories such as blond and black [4]. These variants have then been modeled into accurate prediction systems like HIrisPlex, IrisPlex, The Snipper 2.5 suite and OCME eye color prediction tool [5-9]. The contribution of these markers towards prediction in most cases however, only factors in an independent effect and does not take into account potential interactions or epistasis in the production of the final phenotypic color. Epistasis is a phenomenon that occurs when a gene's effect relies on the presence of other genes, and can display different effects in gene combinations. The current prediction models implement only the markers that have marked independent effect and not their epistatic interactions. Implementing the interactions in the statistical models 'might' improve their prediction accuracy. The HIrisPlex system can predict hair colors with an overall average of 79% (prevalence adjusted accuracy) by providing probabilities of blond, brown, red or black that sum to 1 in its prediction. The average prediction accuracies in HIrisPlex system for blond hair color is 69.5%, brown - 78.5%, red - 80% and black- 87.5%. It is worthwhile to investigate if this

prediction accuracy can be improved, especially in categories like blond and brown where the average accuracies are lower [6], by potentially including epistatic interactions as a new variable in the HIrisPlex system. Age-related darkening of hair color, especially blond hair, needs to be addressed as the current prediction model does not accurately predict the current hair color of an individual whose hair has progressively darkened by age, it predicts their lighter blond hair phenotype[6]. Furthermore, the HIrisPlex system can only be used to predict scalp hair color and not beard hair color. Even if we use the HIrisPlex to predict the beard hair color now, it would not produce an accurate single result, as a lot of individuals, based on their pictures, seem to have an increased number of naturally different beard hair color follicles than their scalp hair color follicles. Some have more than one beard hair color, while their scalp hair is mostly mono-chromatic (apart from greying hair color follicles). Although it is easy to assume that in general, the genes that are responsible for a particular scalp hair color should also be responsible for a male's overall beard hair color, it still remains unclear how certain hair follicles in some individuals display intra-individual color variation. This type of variation even displays the extremes of several hair color defined categories in a single beard growth that is not displayed in their head hair. Growth location and the expression of certain pigmentation associated genes may unearth differences between head and beard hair, particularly beard hair follicles of differing pigment. Essentially, the questions being asked are: can differing hair follicle pigments be attributed to differences in RNA expression patterns, or are there combinations of epistatic interactions that modify the phenotype?

This study, therefore, has two aims to research human pigmentation. First, by looking at epistatic interactions between established pigmentation markers or genes previously associated and found to have predictive value for head hair color, this study's first aim is to check for improvement to existing DNA intelligence tools like HIrisPlex. Particularly, to improve its blond and brown hair color prediction capability. The second goal is to understand differential gene expression patterns between different beard hair colors obtained from the same individual(s). The generation of preliminary results on several

pigmentation candidate genes using RNA expression in differing beard hair follicles may also aid in the understanding of the effects of age-dependent hair color changes - certain individuals, who darken in hair color as they go through adolescence (or even earlier). Beard hair growth of differing hair color follicles may mimic this phenomenon, and also limit inter-individual variation problems associated with RNA expression levels for these candidate pigmentation genes, therefore allowing a unique examination of these pigmentation candidate genes.

1.2 Pigmentation Biology

Pigmentation of eye, hair, and skin and its diverseness amongst different population subgroups has always been an intriguing characteristic. The variety of skin, eye and natural hair colors that we observe today are due to melanin, a pigment synthesized in melanocytes. Melanin is located in our hair bulbs, iris and also in the basal epidermal layer. It is formed from tyrosine through the process of melanogenesis in organelles called melanosomes. *Tyrosinase*, *Tyrosinase related protein 1* are few of the key genes involved in the process[10].

Depending on the presence or absence of cysteine, two different types of melanin are formed (Figure 1.1). Eumelanin is responsible for dark color and is a brownish black pigment, whereas pheomelanin is responsible for lighter color and is a reddish yellow pigment. It is the ratio of eumelanin to pheomelanin that actually results in a particular color. The higher the ratio, the darker the color [11].

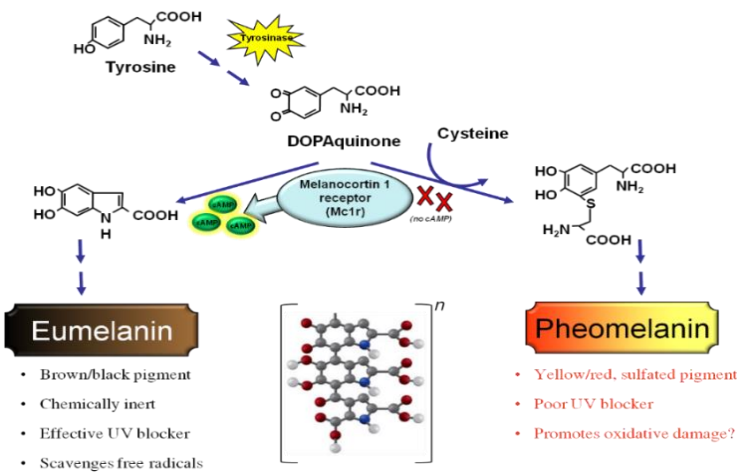

Figure 1.1 Formation of different types of melanin

(This image can be found at *http://doraziolab.uky.edu/)*

The following gives a brief overview of hair follicle growth and the layers of a hair shaft. The cross section of the hair can be divided into cuticle, cortex, and medulla. The cuticle is the outermost layer that reflects the light. The cortex is made up of fibrous cells with proteins. The fibers are surrounded and held together by soft keratins. The cortex also contains melanin, and based on the amount of melanin present, the color of the hair is determined [12]. The melanin produced in melanocytes migrate into the keratinocytes and that gives rise to hair shaft color [13]. The medulla is the center of the hair shaft and its diameter is directly proportional to the thickness of the hair. The cortex is useful in identifying the different hair types, and the cuticle is used to identify the species from which the hair came [12].

Human hair goes through five different cycles,[14, 15]

1. Anagen: Is the active growth phase, which lasts between 2 and 6 years per hair. Approximately, 85% of the hair is in this phase. Apparently, melanogenesis occurs only during this phase.
2. Catagen: Is the transition phase in which the follicle is starting to separate from the dermal papilla. This lasts up to 2 weeks. Melanogenesis gets turned off here.

3. Telogen: Is the resting phase where the follicle completely separates from the dermal papilla and lasts up to 6 weeks. Melanogenesis is completely absent during this phase.
4. Return to Anagen: The hair matrix is formed afresh with the dermal papilla moving back into the hair follicle.
5. Exogen: This is the phase where hairs shed naturally.

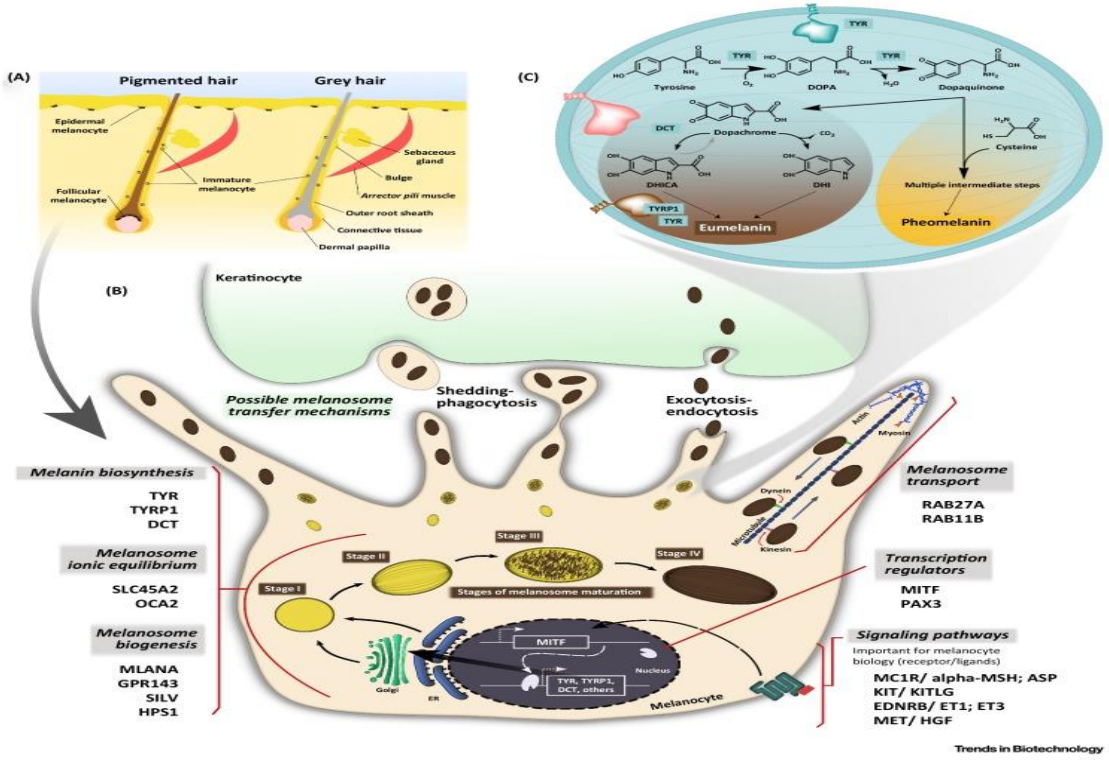

Figure 1.2 Overview of the function of melanocytes and it's migration into keratinocytes in hair follicle pigmentation [16]. A) Diagram of pigmented and grey hair follicle. B) Differentiated melanocytes with frequently associated genes. C) Biochemical synthesis of eu and pheomelanin.

(This image can be found at

http://www.cell.com/trends/biotechnology/fulltext/S01677799(15)00207-3)

1.3 Epistatis Investigation using select pigmentation SNPs

GWAS help in identifying the common variants in a population of individuals possessing a similar trait. Sample sets of cases are compared with sample sets of controls, and the genes or SNPs causing distinctive externally visible characteristics are predicted. Pigmentation is a complex polygenic trait, facilitated by a variety of factors other than the SNPs or variants that have associative or predictive value. One such factor is epistatic or gene-gene interactions. One of the important genes is *OCA2* (located on chromosome 15). Its relationship with pigmentation was deduced by identifying its role in the pigmentation disorder, albinism [17].

As of now, few GWAS focus on gene-gene interactions that attribute to the variance in pigmentation phenotypes. The variants or genes causing significant independent effects are being constantly unearthed. Branicki *et al.,* in 2009, confirmed the association of rs12913832 with eye, skin and hair color by running regression analysis with known SNPs as confounding covariates. They also identified that rs12913832's interaction with *MC1R* seemed to affect skin pigmentation the most [18].

There could be interactions between existing pigmentation genes, which might explain particular pigment variances due to gene combinations. For example, rs12913832 in the *HERC2* gene, has been known to act as a functional regulator, affecting the downstream expression of the *OCA2* gene*,* which has been strongly associated with eye and hair color. The functional presence of allele 'A' in rs12913832 has been shown to enhance the expression of *OCA2*. This leads to an increase in melanin production, thereby leading to darker pigmentation. Its derived allele, 'G' has been shown to have a down-regulatory effect on *OCA2*'s expression levels due to conformational positioning [19].

Various research groups around the world have been trying to understand the differences in pigmentation among different population subgroups. Sturm *et al.,* identified *TYRP1* as having the second strongest association with eye color after *OCA2* [20]. A variant in

TYRP1, rs683, was identified to be associated with iris pigmentation by Frudakis *et al.,* in 2003 [21]. Sulem *et al.,* [22] identified six genes; *MC1R*, *SLC24A5*, *SLC45A2*, *ASIP*, *OCA2,* and *TYR,* which had significant independent effects on pigmentation variation. rs12592730, a part of one of the haplotypes in *HERC2,* was identified to be significantly associated with eye color, in a study that found blue eyes could be due to a perfect association of *HERC2* regulatory element with *OCA2* expression [23]. On the other hand, another study found that, just the combination of rs12913832 (*HERC2*), rs1426654 (*SLC24A5*) and rs16891982 (*SLC45A2*), accounted for approximately 76% of variance in eye color [24]. Eriksson *et al.,* [25] identified rs7183877 (*HERC2*) and rs1667394 (*HERC2*) to be associated with intermediate eye colors like green/ hazel. A patent by Kayser *et al.,* [26] provides a list of SNPs that are correlated or in LD with known variants of human iris colors. There have been a handful of studies performed on known pigmentation markers, to investigate genotype combinations. Ruiz *et al.,* in 2012, performed a study on 416 individuals from 6 different populations, to improve eye color prediction. Their study implemented a non-parametric analysis called multi-factor dimensionality reduction (MDR), which takes in a data set and randomly assigns each sample to training or test set, to calculate the possible combinations of SNPs that are likely to predict a phenotype in question. They identified a significant synergistic interaction with a cross-validation score of 8/10, between rs12913832 and rs1667394 (both in *HERC2*) in green-hazel/ non-green-hazel eyes [7]. Pospiech *et al.,* in 2014 [27], analyzed 38 SNPs in 13 genes, to identify interactions between SNPs that were most correlated to phenotypes, by implementing MDR. This study was able to identify that interactions between rs12913832 (*HERC2*) and r1800407 (*OCA2)* were promising for predicting green-eye perceived color [27]. A South American study in 2014, collected samples from admixed individuals, and assessed the predictive value of different SNPs in their ability to predict green-hazel eyes [28]. They compared six different methods by adding different SNPs. The researchers found out that Snipper13 (one of the methods) appeared to have 50% less chance of error in predicting brown eyes. Rs4778282 (*OCA2*) was one of the SNPs in Snipper13 set [28].

The *MC1R* variants; rs1805006, rs11547464, rs1805007, rs1805008, rs1805009, rs1805005, rs2228479 and rs885479, which are associated with pale skin/red hair traits, comes from knowledge obtained in a twin – familial study conducted by Sturm *et al.*, in 2003 [29]. For studying hair color associated variants, GWAS approaches by Han *et al.*, and Liu *et al*., identified rs12896399 (*SLC24A4)*, rs12821256 (*KITLG*), rs1393350 (*TYR*), rs12203592 (*IRF4*) to be significantly associated with hair color. These studies were amongst the first to identify rs12913832 in *HERC2* (a gene upstream of *OCA2),* and associated it with eye and hair color [30, 31]. Sulem *et al.,* [22] found the variant rs4959270 in *EXOC2,* to be associated with hair color. More variants of *MC1R*; rs1110400 [32], the INDEL N29insA, and Y15OCH were also identified to be associated with a red hair phenotype. A variant in *SLC45A2,* rs28777, was found to show noticeable variance (with AA correlated to black hair and CC to red hair phenotypes). A Scottish-Danish study that incorporated Illumina Microarray and MALDI-TOF technologies for SNP typing, identified the following SNPs, rs16891982 and rs26722 (*MATP*/ *SLC45A22*), rs916977, rs1129038, rs11636232 and rs2238289 (all *HERC2*), rs1470608 (*OCA2*), to be associated with darker hair color, and rs10777129 (*KITLG*) to be associated with lighter hair color [33]. Eriksson *et al.,* in 2010 also identified that- rs12931267 (*FANCA*), rs4778241 (*OCA2*), rs4778138 (*OCA2*), rs7495174 (*OCA2*), influence hair color [25]. Additional hair color SNPs that were included in HIrisPlex prediction model include- rs1042602 (*TYR*), rs2378249 (*PIGU*), and rs683 (*TYRP1*). A study that implemented whole genome sequencing on 2230 Icelanders, found that the variant rs12203592 – T allele in *IRF4* was strongly associated with brown hair, and high skin sensitivity following sun exposure [34]. Another GWAS in an admixed Latin American sub-population attempted to identify SNPs associated with hair greying, hair color, balding, beard thickness, hair shape, etc. This study identified the known variants rs12913832 (*HERC2*), rs12203592 (*IRF4*) (also found to be associated with hair greying) to be associated with hair color variation. They also found rs183671 (*SLC45A2*), rs598952 (*TYR*), and rs1426654 (*SLC24A5),* to be significantly associated with hair color, with p-values $<10^{-5}$ [35].

A number of studies have come out referencing the associations of different variants to skin color. Dr. Murray Brilliant, in a report submitted to the US Department of Justice, has reported that the rs3212355 (*MC1R*) variant was correlated with skin reflectance, although not quite significant [36]. Bonilla *et al.,* identified that the polymorphism of rs6058017 from G to A in the agouti signaling protein (*ASIP*) leads to determining darker to lighter skin color [37]. Lao *et al*., in 2007 looked for signatures of positive selection in genes that were associated with skin pigmentation and performed regression analysis on rs3782974 (*DCT*), rs1800414 (*OCA2*), rs1448484 (*OCA2*), rs2762464 (*TYRP1*) and rs16891982 (*SLC45A2*) within 51 different population subgroups and ascertained that these 5 candidate SNPs explained about 82% of variance in skin pigmentation [38]. A melanoma study reported, rs13289 (*SLC45A2*) and rs1408799 (*TYR*) to be associated with intermediate skin color, and rs1126809 to be associated with skin color in Europeans [39].

A study on association of melanin content with *OCA2* polymorphisms in East Asian population showed that rs1545397 had a drastically different allele frequency between East Asian population and non-East Asian population sub-sets. This SNP is also used to predict lightness/darkness of the skin [40] [41]. A GWAS study by Liu *et al.,* identified *DEF8* to be significantly associated with melanoma. Given that it lies approximately 30Kb downstream of *MC1R* [42], it is probable that an association with pigmentation is also plausible. Jacob *et al.,* [43] followed a candidate gene approach to skin color association using digital quantification of imagery based on Hue- Saturation Value (HSV) color space to identify continuous skin color variations from 14,185 SNPs (281 genes). They identified rs10756819 (*BNC2*) and rs1712891 (*SLC24A4*) as having a potential effect on skin pigmentation prediction [43]. Maronas *et al*., in 2014, implemented naïve Bayesian analysis, MDR, and principal component analyses, to identify associations to skin color from a set of 59 SNPs. All the above techniques are used to reduce the number of dimensions into understandable components. They selected these 59 SNPs based on prior literature knowledge and information available from the 1000genomes database. This study resulted in developing a predictive model for skin color with 10 SNPs (8 genes) [44],

for which they achieved relatively high levels of prediction accuracy: White (AUC= 0.999), intermediate (AUC= 0.803), and black (AUC=0.966). However, due to their small sample set of approximately 200 individuals, this model still needs to be validated on an increased global set of individuals to truly test its performance. In 2015, Jacobs et al., calculated an R^2 change to explain the variance in skin color phenotype within their model and found that rs6059655_A (*RALY*) caused a significant effect on skin color prediction [45].

As seen above, almost all the pigmentation genes have an overlap and several have predictive capabilities for pigmentation variation. Although all these studies are identifying independent and interactive effects of SNPs, a lot of additional analyses need to be performed in order to increase the accuracy of predictions. The accuracy for prediction of blond hair in HIrisPlex for example, is only about 69.5% and so more research needs to be performed to increase these levels of accuracy. Interactions between genes that were associated with different traits might have been missed out, as the power of association was not strong enough for an independent effect. Looking into a more candidate gene approach, it may unearth combination effects on the final phenotype. Therefore, it is best to investigate all known predictive pigment genes or markers. It is our belief that searching for epistatic interactions would be a major step forward towards understanding potential genotype patterns/combinations that may enhance current prediction accuracy levels.

1.4 Pigmentation and Gene Expression

Androgen-based differences in genes were studied between scalp and beard hair [46]. The study showed that genes *SFRP2, MN1, ATP1B1* and *Fibulin 1D* were highly expressed in beard hair follicles versus scalp hairs. *TGF-B2* was also highly expressed in scalp hair versus beard hair [46]. Pielberg *et al.,* studied premature hair graying in horses and associated it with a duplication in *STX17* gene's intron 6. An increase in expression for *STX17* and *NR4A3* have been found in gray horses [47]. A study performed by Chiaverini *et al.,* reported that low levels of the gene *CTNS* could be the cause for autosomal

recessive disorder: Cystinosis. The silencing of the *CTNS* gene causes 75% reduction in the melanin synthesis mainly because, low levels or absence of the *CTNS* gene results in the degradation of *TYR* (a critical gene in the pigmentation pathway) [48]. A study [49] on identifying genes that were differentially expressed in gray, white and pigmented human hair follicles identified that genes such as *TYR, TYRP1, PMEL, MLANA, SLC45A2, PLXNC1, GPR143*, etc., were downregulated in both gray and white hair follicles when compared to pigmented hair follicles. Korberg *et al*., [50] identified that repeat polymorphism in *MITF-M* promoter regions in dogs were responsible for reduced white spotting [50]. Studies on mice models have shown that the expression of *KITLG* (ligand for *KIT*, a tyrosine kinase receptor) is also down-regulated in lighter pigmented hairs [51]. One study attempted to identify differentially expressed miRNA between different coat (black and white) follicles of sheep. The list of differentially expressed MiRNAs identified were validated using qPCR analysis [52].

The variant rs12203592, is a well-known pigmentation SNP in IRF4 that modulates the expression of *IRF4* gene- Lower expression of which, has been associated with lighter hair color [34, 35, 53]. miRNA 214 was identified to modulate the Wnt pathway thereby controlling skin and hair follicle development. An increase in *miRNA* 214 was found to have decreased the expression of *β- catenin* and *Lef-1* and therefore reducing the bulb size and producing thinner hair [54]. Expression patterns of the *SLC7A11* gene (a critical regulator of the pheomelanin production in melanocytes and hair) was assessed in alpacas using sequencing and Real Time PCR data. This showed that high levels of *SLC7A11*'s expression was associated with brown coat color in alpacas [55]. Since this gene has not been reported frequently in human melanogenesis pathway, it could be a potential target in future studies on beard hair pigment genes.

Overall, there are many differentially expressed genes between differing hair color follicles. Therefore, it is important to understand the expression pattern of these genes and other candidate genes in order to understand how changes in expression can be

related to genomic sequences. This may enhance understanding of pigmentation pathways and help provide a stage for improved statistical modeling tools for phenotype prediction purposes.

CHAPTER 2. MATERIALS & METHODS: EPISTASIS

2.1 Sample Collection

Samples were collected as a part of IRB #1409306349 (n = 546) and samples that were a part of an Irish study (n=346) were made available for this project (all with ethical consent). 5ml of saliva sample was obtained from each volunteer. At the time of the sample collection, the volunteers were digitally photographed to capture their pigmentation and were asked to fill out a survey. Nikon camera with AF-S Micro NIKKOR 60mm lens and Nikon WIRELESS SPEEDLIGHT SU-800- REMOTE SB-R200 flashes were used for photographing. A calibration picture was taken prior to each sample collection using SpyderCHECKR^{TM} to ensure the settings and lighting conditions matched across different collection periods. Apart from the pictures, spectrophotometric wavelengths measuring reflectance at various wavelengths (400-700nm), melanin index, l, a, b values, etc., of hair and skin color were obtained using Konika Minolta's Canon 700d/600d spectrophotometer.

2.2 DNA Extraction, Quantitation and SNP Genotyping

DNA of 546 US samples were extracted following an in- house salting out method. 1ml of saliva sample mixed with lysis buffer, 15ul of Proteinase K and 130ul of 10% SDS (Dot Scientific Inc., Burton, MI) were incubated in 2ml tubes in a thermos block, set to 300 rpm at 37°C overnight. Post lysis incubation, 200ul of 0.5M NaCl (Dot Scientific Inc.) was added and then incubated at room temperature for 10 minutes. After which, the samples were centrifuged for 12 minutes. The supernatant was then transferred to a new tube and an equal volume of 100% isopropanol (Fisher Scientific International Inc., Hampton, NH) was added and incubated for 10 minutes at room temperature.

Following the incubation, the samples were centrifuged for 15 minutes. The supernatant was then discarded and the pellets were washed twice with 70% ethanol (Decon Laboratories Inc., King of Prussia, PA) and centrifuged for 5 minutes after the first wash. Post second wash, the supernatant was carefully removed and the pellets were air- dried for 10 minutes in the hood at room temperature. The samples were then re-suspended in 30ul of DI H_2O and stored at -20°C or -80°C until further use. All the centrifugations were done at 16200cfg and at 4 to 10°C. All the samples were quantified using Qubit® DNA HS Assay Kit used with Qubit® fluorometer. The 346 Irish samples were previously extracted but were newly quantified to give a more accurate concentration reading.

SNP Genotyping:

In total, 61 SNPs found in 19 known pigmentation genes were chosen for this study to understand epistatic interactions or associations that may lead to color modifications. SNP Multiplex Genotyping was carried out on these 61 SNPs by splitting them into four different assays for the US samples. The assays are HIrisPlex (24 SNPs)[6, 56], HIrisPlex-S (17 SNPs), ComPlex (13 SNPs)[7] & Philips-Skin (7 SNPs)[44]. SNP breakdown per assay is listed in Appendix A.

Pre-designed flanking primers and single base extension (SBE) primers that were available in-house were used for HIrisPlex and HIrisPlex-S assays. The flanking primers and SBE primers for Philips- Skin assay and ComPlex assay were designed in house. For designing flanking primers, flanking sequences for each SNP was obtained from dbSNP [57]. The flanking sequence was then plugged in to Primer3Plus[58] and the program was trained to exclude neighboring SNPs from the primer design. Information about neighboring SNPs was obtained from dbSNP. A quick ePCR was performed using BiSearch[59] tool to check for the specificity of the primer pair. Once primers for all the SNPs were designed, all the primers were loaded onto Autodimer[60] program to see if any of the primers form primer-dimers with each other. BatchPrimer3 was used to design Single Base Extension primers. Once SBE primers for all the SNPs were designed, they were checked to rule out

any potential primer-dimers using Autodimer program. All the primers are listed under Appendix A.

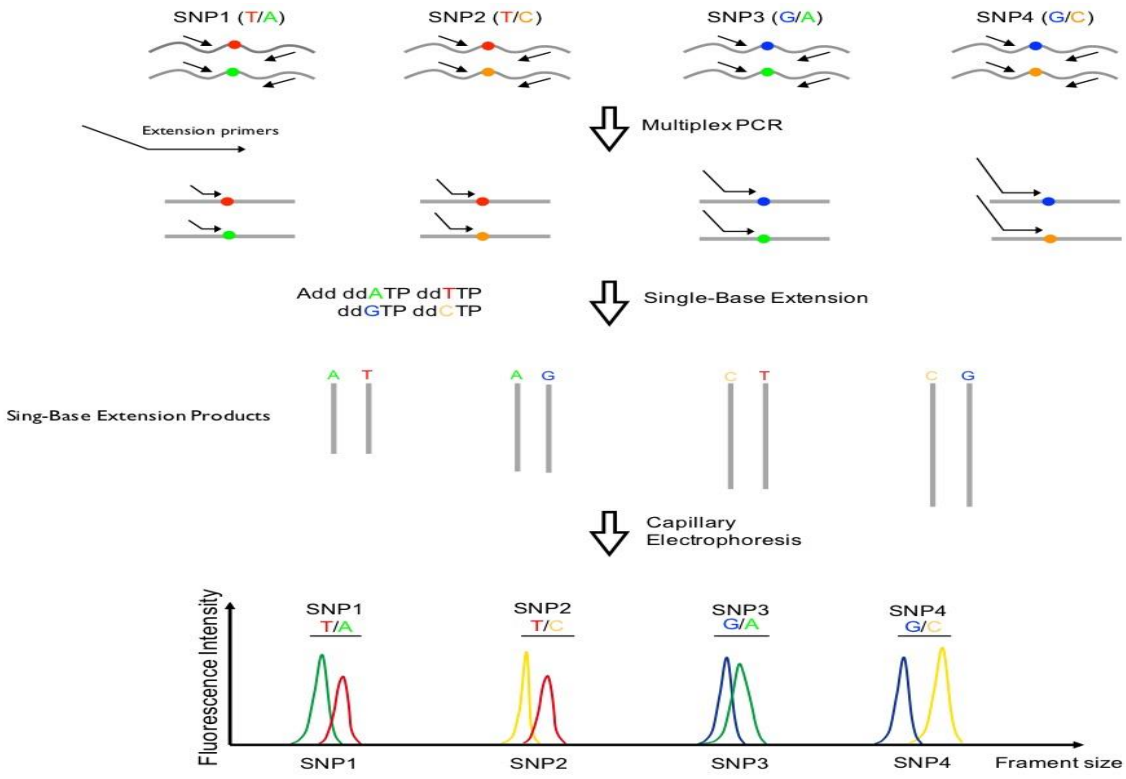

Figure 2.1 Overview of SNaPshot multiplex SBE assay

(Image found at http://biotech.geneskies.com/en/index.php/Index/fuwuer/id/27)

The SNPs of interest were amplified following a multi-step, single base extension (SBE) method of PCR. For the first step, 80- 250bp regions containing the SNP of interest were multiplexed and first amplified using 0.4uM of forward and reverse primers. Approximately, 1ng of DNA was amplified in a total reaction volume of 10uL. 1ul of 10X PCR gold buffer, 1uL of 25uM $MgCl_2$, 0.22uL of 10mM dNTPs and 0.3uL of 5U/uL Taq Gold (Life Technologies Corporation, Carlsbad, CA) were used to amplify the regions of interest. PCR amplification was performed on Eppendorf Mastercyclers- Nexus SX1 or GX2 using the following set up: 95°C/ 10 min; [95°C/ 30s; 61°C/ 30s] $_{33}$; 61°C/ 5 min. Then the PCR products were cleaned using EXOsap- IT (Affymetrix, Inc. Santa Clara, CA) and the protocol specified on the product was followed. Once purified, SNaPshot multiplex assay was

conducted on the PCR products by adding SBE primers. 2uL of purified PCR products and 1uL of SNaPshot reaction mix was used for each sample. The following thermal cycling set was used for SBE reactions: 96°C / 2 min; [96°C/ 10s; 50°C/ 5s] $_{25}$; 60°C/30s. The final SBE PCR products were purified using Shrimp Alkaline Phosphatase (Affymetrix, Inc.).

The purified SBE products were then separated and detected using capillary electrophoresis on Life technologies' ABI 3500 Genetic Analyzer (POP7 and 50cm array). 1uL of purified product was used for electrophoresis and SNaPshot multiplex protocol was followed. The data obtained was analyzed using GeneMapper software (Applied Biosystems). The minimum peak threshold was set to approximately 50 RFU to ensure correct genotyping.

For the Irish samples, genotypic information for 38 SNPs were extracted from the Identitas study [61]using the extract command on PLINK program [62]. 17 other SNPs were already genotyped and available in the database. The remaining 6 SNPs were multiplexed as a separate assay and all 346 samples were genotyped using this SNaPshot SBE assay.

2.3 Phenotyping

The samples were phenotyped for hair color by three individuals. All three individuals viewed the images with the same brightness/contrast setting on an Apple iMac and a color reference guide was used to maintain consistency. These were compared to the self-reported information for validation check. For individuals that were reported to have dyed their hair, self-reported phenotype (what they remember they had in their early 20's) was used. Hair Color was classified as Black, Browns (Dark Brown & Light Brown), Reds (Red/Auburn, Pure Red and Light Red/Strawberry Blond) and Blond.

To display hair color wavelengths of the categorically phenotyped individuals, the subjective hair color categories were graphed against the objective assessments obtained

using reflectometry. The average spectrophotometric measurements of all the individuals that did not dye their hair (n= 348) were plotted against the classified hair color phenotypes using Excel 2013 (Microsoft) (Figure 2.2).

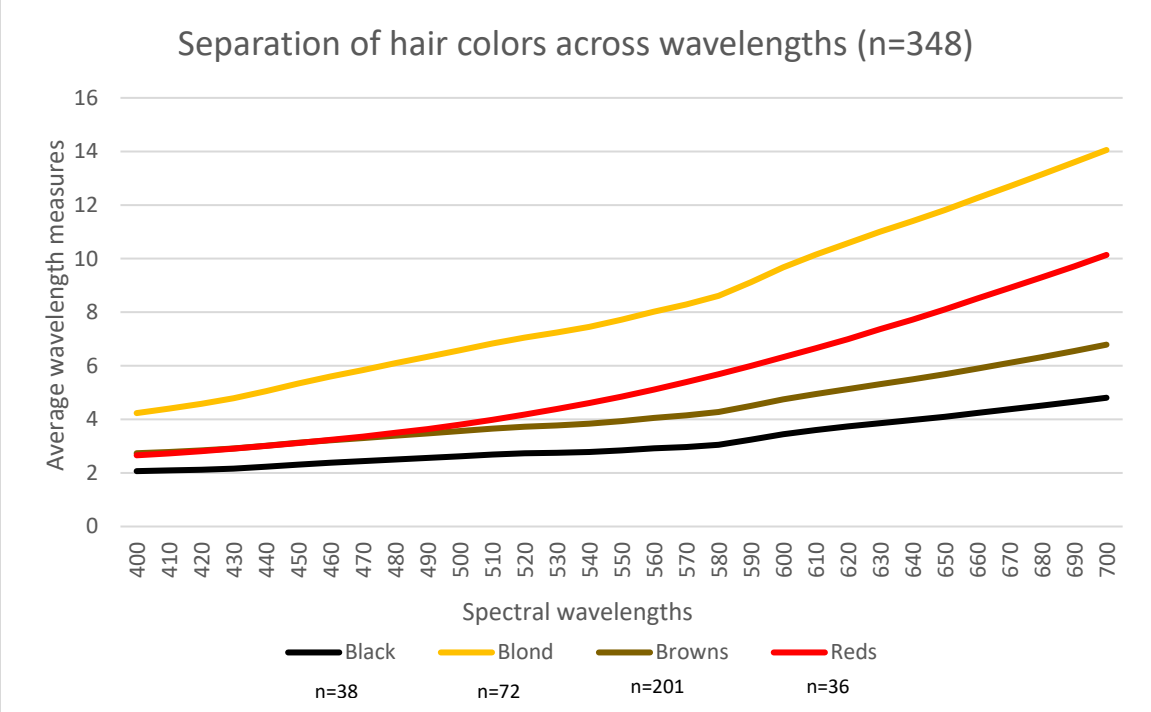

Figure 2.2 Separation of hair colors across wavelengths (400-700nm)

To show individual variation within the dataset, the means of 700nm wavelength with standard deviation was plotted (Figure 2.3). It is suggested by Frudakis in 2010[63] that 650nm is the best wavelength to notice separation of hair colors. However, we found that the higher level measure better separated the categories upon visual observation and measure within our dataset.

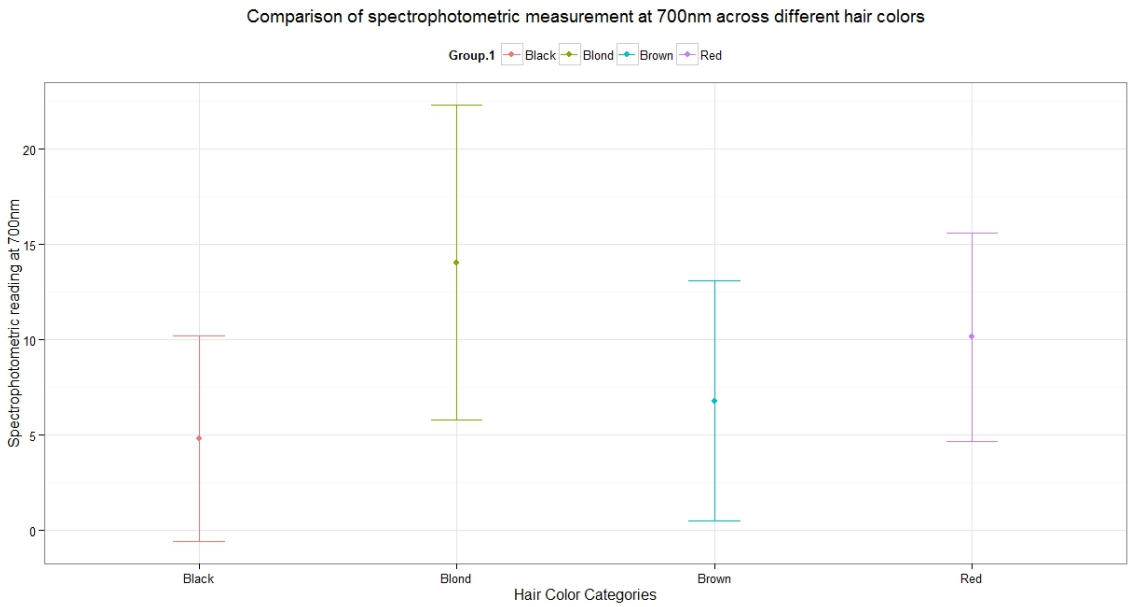

Figure 2.3 Plot of means at 700nm for each hair color category with SD bars

2.4 Statistical Analysis

Epistatic interactions are measured by examining combinations of genotypes between SNPs. To begin, all single markers were binary coded with minor allele as target. If the genotypes were homozygous with just the minor allele, then the genotypes were coded as '2'. If the genotypes were heterozygous with one minor allele, they were coded as '1' and if the genotypes were homozygous in the absence of minor allele; they were coded as '0'. This binary code is then multiplied by the two SNPs being investigated and the effect is measured as an interaction (bioinformatically and statistically) due to the presence/ absence of these minor allelic combinations.

For this study, 61 SNPs were tested for potential gene-gene interactions which might add on to the explanation of variation in human hair color. Minor Allele Frequency (MAF) for each of the SNPs were calculated using PLINK program (Table 2.1). As expected, the *MC1R* variants-rs3212355, Y15OCH, N29insA and rs11547464 had MAF < 0.01. Three of these are important pigmentation markers especially for red hair and they are known to occur

at lower frequencies [64]. Since the end goal was to compare the new model to existing prediction models, using the SNPs that are already present in the existing models was important, therefore they were kept in for analyses.

Table 2.1 Minor Allele Frequency table for 61 markers. Highlighted in 'bold' are MAF <0.01

SNP	Gene	Reporting publication	Chromosome	Chromosomal location	Allele variants	MAF
rs13289	*SLC45A2*	[39]	5	33986304	G/C	0.3953
rs16891982	*SLC45A2*	[33, 41, 65]	5	33987450	C/G	0.1299
rs26722	*SLC45A2*	[33]	5	33963765	T/C	0.03261
rs28777	*SLC45A2*	[29]	5	33994716	C/A	0.1093
rs4959270	*EXOC2*	[22]	6	402748	A/C	0.4994
rs12203592	*IRF4*	[30, 31] [34]	6	341321	T/C	0.2569
rs10756819	*BNC2*	[43]	9	16858086	G/A	0.3173
rs1408799	*TYRP1*	[39]	9	12662097	T/C	0.3668
rs683	*TYRP1*	[21]	9	12699305	C/A	0.4077
rs35264875	*TPCN2*	[66]	11	68602975	A/T	0.4001
rs3829241	*TPCN2*	[66]	11	68611939	A/G	0.3592
rs1042602	*TYR*	[67]	11	88551344	A/C	0.3051
rs1126809	*TYR*	[39]	11	89284793	A/G	0.2415
rs1393350	*TYR*	[30, 31] [34]	11	88650694	A/G	0.2822
rs10777129	*KITLG*	[33]	12	88567936	A/G	0.1139
rs12821256	*KITLG*	[30, 31] [34]	12	87852466	C/T	0.09714
rs12896399	*SLC24A4*	[30, 31] [34]	14	91843416	T/G	0.3931
rs17128291	*SLC24A4*	[45, 68]	14	92416482	C/T	0.1613
rs2402130	*SLC24A4*	[22]	14	91870956	G/A	0.2265
rs1129038	*HERC2*	[33]	15	28111713	G/A	0.3184
rs11636232	*HERC2*	[33]	15	28141480	T/C	0.3555

Table 2.1 continued

rs12592730	*HERC2*	[69]	15	28285213	A/G	0.05805
rs12913832	*HERC2*	[30, 31] [34]	15	26039213	T/C	0.3184
rs1667394	*HERC2*	[25]	15	28285036	G/A	0.2226
rs2238289	*HERC2*	[33]	15	28208069	C/T	0.1871
rs6497292	*HERC2*	[69]	15	28251049	C/T	0.08183
rs7183877	*HERC2*	[25]	15	28120587	A/C	0.1064
rs916977	*HERC2*	[33]	15	28268218	A/G	0.2124
rs12441727	*OCA2*	[26]	15	28026629	A/G	0.1594
rs1375164	*OCA2*	[70]	15	28046666	T/C	0.2382
rs1448484	*OCA2*	[38]	15	28038295	C/T	0.05063
rs1470608	*OCA2*	[33]	15	28042975	A/C	0.1699
rs1545397	*OCA2*	[41]	15	27942626	T/A	0.06671
rs1800407	*OCA2*	[18]	15	25903913	A/G	0.1019
rs1800414	*OCA2*	[38]	15	27951891	C/T	0.01683
rs4778138	*OCA2*	[70] [25]	15	28090674	G/A	0.158
rs4778232	*OCA2*	[69]	15	28036619	T/C	0.2333
rs4778241	*OCA2*	[25]	15	28093567	A/C	0.2126
rs7495174	*OCA2*	[25]	15	28099092	G/A	0.08228
rs8024968	*OCA2*	[69]	15	28038543	T/C	0.12
rs1426654	*SLC24A5*	[71]	15	48134287	G/A	0.07323
rs3114908	*ANKRD11*		16	89317317	A/G	0.353
rs8051733	*DEF8*	[72]	16	89957798	C/T	0.4446
rs12931267	*FANCA*	[19]	16	89752324	G/C	0.0908
rs1110400	*MC1R*	[29, 73]	16	89986130	C/T	0.01076
rs11547464	*MC1R*		16	89986091	A/G	**0.007361**
rs1805005	*MC1R*		16	89985844	T/G	0.1315
rs1805006	*MC1R*		16	89985918	A/C	0.01529
rs1805007	*MC1R*		16	89986117	T/C	0.09668

Table 2.1 continued

rs1805008	*MC1R*	[29, 73]	16	89986144	T/C	0.06519
rs1805009	*MC1R*		16	89986546	C/G	0.01529
rs2228479	*MC1R*		16	89985940	A/G	0.1059
rs3212355	*MC1R*		16	89917970	A/G	**0.000608**
rs86insA	*MC1R*		16	89985753	A/C	**0.001133**
rs885479	*MC1R*		16	89986154	A/G	0.05778
Y152OCH	*MC1R*		16	89986122	A/C	**0.001133**
rs1015362	*ASIP*	[66]	20	32202273	T/C	0.2823
rs6058017	*ASIP*	[74]	20	34269192	G/A	0.1039
rs6119471	*ASIP*	[41]	20	34197406	C/G	0.02673
rs2378249	*PIGU*	[75]	20	32681751	G/A	0.1778
rs6059655	*RALY*	[76]	20	34077942	A/G	0.09307

Since, there were not a considerable number of individuals in the sub- groups (red-auburn, strawberry blond/light red and pure red) of red hair color, all the sub-groups were merged into one higher order group – red. For the purpose of extracting maximum information, hair colors were categorized into 5 different categories- black, dark brown, light brown, red and blond. The 5 hair color categories were tested for 10 different classifications of cases and controls. The table below indicates the list of classifications that were tested.

Table 2.2 List of cases vs controls classifications used for MDR and BLR analysis

Classifications tested for MDR and BLR analysis	
Black Vs All	Light Brown Vs All
Black Vs Dark Brown	Light Brown Vs Red
Dark Brown Vs All	Red Vs All
Dark Brown Vs Light Brown	Red Vs Blond
Blond Vs All	Blond Vs Light Brown

2.4.1 Multifactor Dimensionality Reduction

The interactions between the 61 known pigmentation markers were tested using multifactor dimensionality reduction approach (MDR; software ver. 3.0.2 (www.epistasis.org)). The method, as the name suggests, reduces high dimensional data by grouping the genotypes into high or low risk based on the ratio of cases to controls. The phenotypes were binary coded, '0' being the controls, and '1' being the cases. The method follows 10 fold cross validation statistics, wherein the data is divided into 10 equal parts. 9 parts are considered as training sets, and 1 part is considered as the test set. The models are generated based on the training set, and then are tested for accuracy on the test set. The validations are performed 10 times, and a mean accuracy for each model is obtained. Since the cases and controls are not equally balanced, another measure called testing balance accuracy is used to gauge how many instances the model correctly classified. Testing balanced accuracy provides equal weight to the classification avoiding any potential bias caused by larger class. Balanced accuracy is calculated as (Sensitivity + Specificity)/2 [where, Sensitivity= true positives/ (true positives + false negatives), Specificity= true negatives/ (false positives + true negatives)] [77].

Prior to MDR analysis of each classification, pre-process filtering was performed using ReliefF filter, in order to reduce false positives. ReliefF statistic checks the relevance of individual markers and the dependency between the markers, in order to predict a phenotype[78, 79]. The number of variables for further consideration were limited by applying this filter. Top 5 and top 20 attributes were selected for further analysis. Since this was reducing the number of attributes significantly, additional filtering methods were also applied and the results from all the analyses were compiled and assessed for further analysis. Other than the ReliefF 5 and 20 filters, MDR analysis was performed under the following conditions:

1. With all the SNPs in LD removed: This way, the SNPs not in LD could show up potential interactions.

2. Without rs12913832 as it is a single most contributing SNP in pigmentation. By removing this SNP from the dataset, interactions that are potentially hidden could show up.
3. Without rs16891982 as it is an AIM SNP. By removing this SNP, interactions that were previously masked due to the effect of ancestry on the phenotype were assessed.

In all the analyses, one to five attribute combinations were applied, and 'sex' was set as a covariate. The statistical significance for the models was assessed by performing 1000-fold permutation analysis using MDR Permutation Testing Module 1.0 beta 2 (www.epistasis.org).

The interactions were assessed by viewing the circle graphs for each analysis. The circle graphs provide the strength of the interactions, and explain the nature of the interactions by applying entropy based concepts of information gain[80]. A negative entropy between two SNPs implicates 'redundancy', meaning, the interaction gives redundant information as that of the individual SNPs. A positive entropy suggests 'synergy' between two SNPs, meaning the information provided by the interaction is much higher than the sum of individual SNPs. Since it is based on entropy, in theory, it means that the entropy value eliminates a certain percentage of 'uncertainty' in the prediction of a particular phenotype (prediction of 'cases' in a case/control test model) [27].

2.4.2 Binary Logistic Regression

For the second part of epistasis interaction assessment, binary logistic regression was implemented. It is a statistical method that estimates the probabilities of success, based on the explanatory variables. As the name suggests, in binary logistic regression the phenotype is binary coded into cases and controls. Datasets with categorical genotypes-phenotypes usually do not satisfy the basic assumptions such as normal distribution, homogeneity of variance, and linear relationship. In these scenarios, logistic regressions

or any non-parametric factor reduction methods are appropriate. A binary logistic regression model uses maximum likelihood estimation instead of ordinary least squares method. Depending on the type of phenotype, PLINK epistasis [62] uses either binary logistic regression or linear regression to calculate pair-wise epistatic interactions between SNPs in an additive manner. The basic formula used by PLINK epistasis is:

$$Y \sim b_0 + b_{1.A} + b_{2.B} + b_{3.AB} + e$$

Where, b_0 is the intercept, $b_{1.A}$ is the regression co-efficient for allele A, $b_{2.B}$ is the regression co-efficient for allele B and $b_{3.AB}$ is the regression co-efficient for interaction AB. Command "—epistasis" was used to test pair-wise interactions. Since, this would result in thousands of pair-wise comparisons, the output was pre-filtered to contain only the interactions that were statistically significant at *p*-value < 0.05. The filtered output contained information about SNP- SNP interactions, their corresponding odds ratio, Chi-square values, and *p*-values.

2.4.3 Model Validation

In order to assess the significance and contribution of each interaction towards the prediction of particular phenotypes, it is important to validate the addition of each interaction into the existing HIrisPlex or IrisPlex models. For the purpose of model validation, multinomial logistic regression (MLR) was implemented in order to gauge the accuracy of prediction. Since, the HIrisPlex model uses 4 categories of hair color [81], dark browns and light browns were grouped into one category called 'Brown'. From here on the MLR and AUC calculations would be performed on 4 hair color categories- black, brown, red and blond. All the MLR calculations were performed using RStudio statistical tool[82] using the formula that was previously published in studies by Walsh *et al.* and Liu *et al.*[5, 31]. mlogit[83] and ROCR[84] packages were used to generate logits and AUC values.

For the purpose of making diplotypes of the interactions, each SNP-SNP combination was grouped into a single variable by recoding in a multiplicative format. For example, the diplotypes were coded as '0' if even one of the SNP has no minor allele in its genotype. It was coded '1', if both the SNPs had at least one minor allele in the genotype. Combinations in which one of the SNP has two minor alleles whereas the other SNP has only one minor allele, were coded as '2'. Finally, if both the SNPs have two minor alleles, they were coded as '4'.

The following template models were tested to assess the contribution of SNPs or interactions to the HIrisPlex model using MLR-AUC method.

1. HIrisPlex+ Interactions within SNPs in HIrisPlex model
2. HIrisPlex+ SNP1
3. HIrisPlex+ SNP1+ Interaction between SNP1 and one of the SNPs in HIrisPlex model
4. HIrisPlex+ SNP1+ SNP2
5. HIrisPlex+ SNP1+ SNP2+ Interaction between SNP1 and SNP2

AUC values were calculated for each of the combinations possible with all the samples in the dataset. Models that increased the AUC of any of the hair color category/ overall accuracy were considered to be better models than HIrisPlex model. However, since 100% of the data is being used, there is a possibility for false positives. In order to assess the model's fitness, they were subjected to cross-validations. Cross-validations are a sophisticated extension of the split-sample approach. The dataset is randomly split into training and test sets based on the split percentage. MLR and AUC calculations are performed on training sets and the probabilities for test samples based on the trained samples are estimated. This process is repeated 'n' number of times based on user's discretion. For our study, we used 90/10 random split, meaning, at a given point 90% of the samples are classified as training samples and 10% of the samples are used to test the prediction.

CHAPTER 3. MATERIALS & METHODS: HAIR FOLLICLE PIGMENT GENE EXPRESSION

3.1 Sample Collection

Samples were collected from males as a part of IRB # 1505822926. Primary inclusion criteria for this study were: the participants must be male of European ancestry with multi- colored beard (natural colors >= 2) and must be 18 years or older. A pre- screening questionnaire determined if an individual qualified for the study. If the individual qualified for the study, he was provided a randomized number, 5 -10 scalp hair were plucked and 5 hairs of each beard hair color were plucked. The hair follicles were immediately submerged in a collection tube containing RNA*later™* solution (Invitrogen*™,* Carlsbad, CA). A survey pertaining to scalp hair and beard hair was filled in by the participants. Photos and spectrophotometric measurements of face, beard and scalp hair were taken using Nikon camera and Canon 600d/700d spectrophotometer.

3.2 RNA Extraction and Quantitation

For the purpose of this pilot study on identifying differential gene expression, 2 individuals that had three different beard hair colors (dark brown, red and blond) were chosen. Different kits were tested and assessed for their ability to isolate RNA from 3-4 hairs. GeneJET RNA purification kit (ThermoFisher Scientific, Waltham, MA), High Pure RNA Isolation kit (Roche Holding AG, Basel, Switzerland), Direct-zol™ RNA Kit (Zymo Research, Irvine, CA) and RNeasy® Micro Kit (Qiagen, Hilden, Germany) were tested. Of the four kits, RNeasy® Micro Kit was the only one that gave quantifiable yield. RNA extraction was performed using the RNeasy® Micro Kit and the protocol was modified in the following manner to fit our requirements:

For each hair color RNA extraction, 3-4 hairs (with follicles attached) were centrifuged for 30s and then RNA*later™* was removed carefully. 350uL of Buffer RLT along with 5uL of Proteinase K was added to the tube and vortexed for 15s. Then, the hairs were ground in the buffer mix using an UV sterilized glass rod for about a minute, followed by pulse vortexing for 30s. The samples were then centrifuged for 3 minutes at 10,000 rpm. The above method of lysis was deemed fit for hair RNA extraction when dealing with smaller number of hairs. In order to avoid/minimize the risk of co-isolating genomic DNA, DNase (Qiagen) digestion step was performed as recommended in the RNeasy® Micro Kit. Finally, RNA was eluted in 17uL of RNase-free water that was provided in the RNeasy® Micro Kit. All the centrifugations were performed at 10,000 rpm at 15°C. The extracted samples were then quantified using Qubit® RNA HS Assay Kit used with Qubit® fluorometer. In order to avoid any stochastic difference due to varied concentrations across the extracted RNA samples, the samples were all diluted to measure approximately 5ng/uL.

3.3 cDNA Synthesis

First strand cDNA synthesis was performed on the extracted RNA using qScript™ XLT cDNA SuperMix (Quanta BIOSCIENCES™, Beverly, MA) according to the manufacturer's protocol. 4ul of RNA template was used for each synthesis and the following amplification conditions were used: (25°C/ 5 min; 42°C/ 60 min; 85°C/ 5 min) on Eppendorf Mastercycler Nexus GX2.

3.4 Quantitative PCR (qPCR) Reaction Set up

cDNA templates were then amplified using RT-qPCR using ABI 7500 (Life Technologies). A probe based Real Time PCR was performed using Taqman chemistry. 32 genes were included in the study based on bioinformatics analysis on another pigmentation study in the laboratory on non-human primates (unpublished). Therefore PCR primers and probes were designed based on conserved regions in non-human primates and humans. The house-keeping gene used was Tata-box Binding Protein (TBP) and it was multiplexed with every gene in order to ensure the presence of cDNA in the samples. Each reaction was

performed with a total volume of 10ul, using TaqMan® universal PCR Master Mix (Thermo Fisher Scientific) with 1uL of cDNA sample, and 750nM of primers and 250nM of probes. In order to ensure minimal competition between primers of gene of interest and housekeeping gene, all the primers used were at the same concentration and volume. The PCR amplification conditions are as follows: 50°C/ 2 min; 95°C/ 10 min; [95°C/ 15s; 60°C/ 1 min] $_{40.}$ Two individuals were analyzed for gene expression. Different hair colors (dark brown, red and blond) were run in duplicates with TBP as endogenous control. Each run contained a negative control (no template) for every gene that was amplified.

3.5 qPCR Data Analysis

The qPCR data analysis was performed using a standard approach where the fold expression is analyzed using $2^{-\Delta\Delta C_T}$ [85]. C_T is the threshold cycle number and is normalized from the C_T of housekeeping/ endogenous control: $\Delta C_T = C_{T\ (Target\ gene)} - C_{T\ (TBP)}$. The next step is to calculate $\Delta\Delta C_T$ which is calculated using: $\Delta\Delta C_T = \Delta C_{T\ (Blond/Red)} - \Delta C_{T\ (Dark\ brown)}$. Dark brown was set as the reference color following Bradley *et al.* [86]. Reduced expression in comparison to reference sample is hard to identify when $2^{-\Delta\Delta C_T}$ is <1 [87]. Hence, a fold change for these samples were calculated using the formula $-1/2^{-\Delta\Delta C_{T.}}$ Single factor analysis of variance was performed using Excel 2013 (Microsoft) for assessing the statistical significance of log fold change.

CHAPTER 4. RESULTS & DISCUSSION- EPISTASIS

4.1 Multifactor Dimensionality Reduction

MDR analysis implements the concept of information gain based on entropy. Entropy represents chaos or disorderliness. MDR analyzes the information gained for each SNP first. In other words, it analyzes the randomness or chaos that a particular SNP can take away from the model while predicting a case/ control. Once this information is obtained for each SNP, it then analyzes SNP- SNP combinations. When analyzing, it looks for whether this combination results in gain or loss of information when compared to the individual SNPs in the interaction. If the combination results in a positive number, it means that the interaction resulted in gain of information when compared to the individual SNPs in the interaction. This type of positive entropy is called a 'synergistic interaction'. However, if the combination gives a negative number, it means that there is information loss and this could be due to high linkage disequilibrium or correlation between the SNPs. These type of negative entropies are called 'redundant interactions'. When there is absolutely no gain or loss of information, then the SNPs in the SNP- SNP combination have an independent effect than an epistatic effect. For the purpose of this study, we focused on synergistic interactions as they give us information about which combination or interaction holds more value than the individual contributors. Figures: 4.1 to 4.10 contain all the circle graphs that we obtained from different MDR analysis. Positive entropy is represented in red and orange lines and they indicate higher and lower degrees of synergistic interaction respectively, blue and green lines indicate higher and lower degrees of redundant interactions and gold line indicates additivity or independence [88]. Results are highlighted in figure legends (Figures 4.1- 4.10).

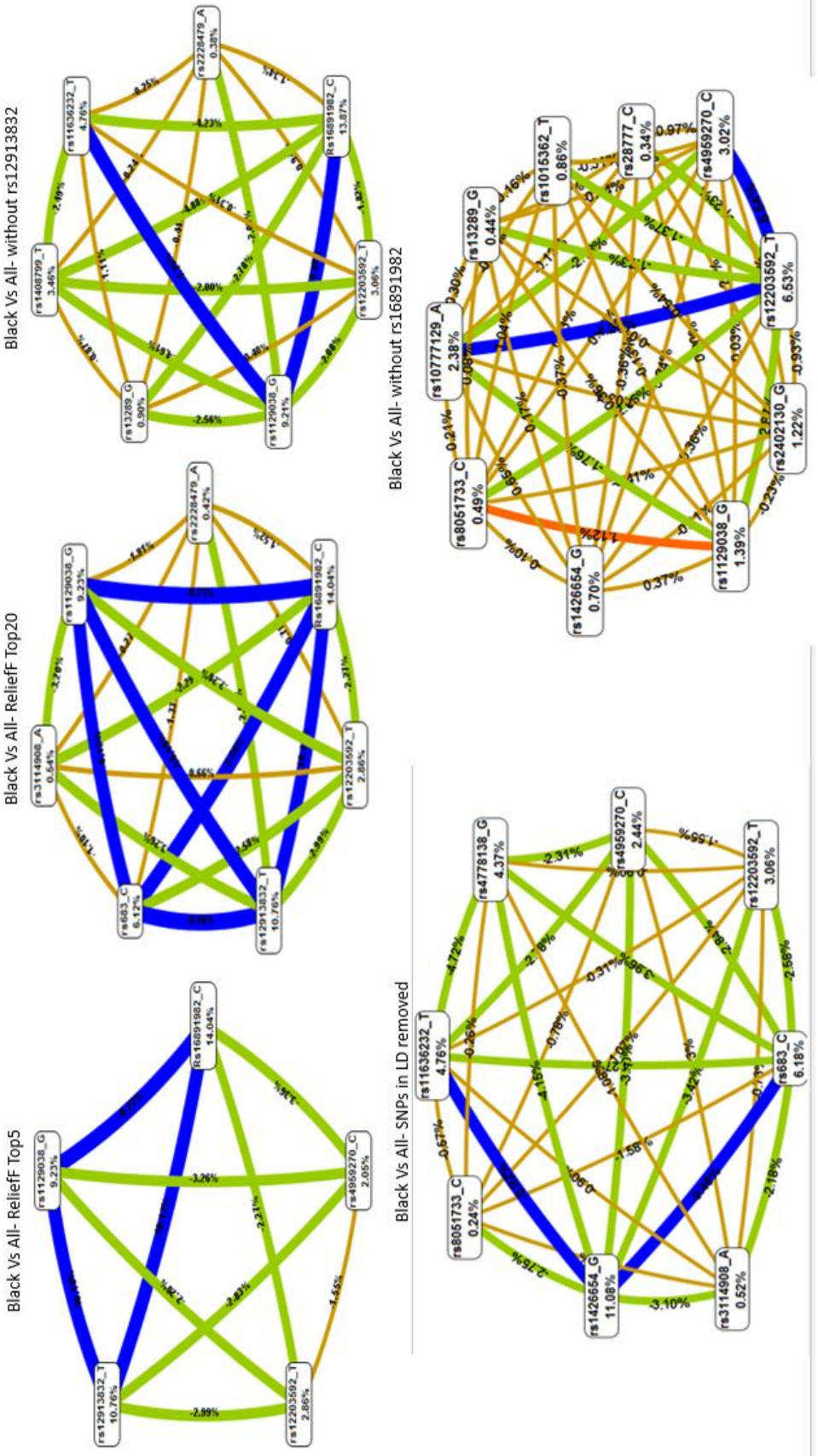

Figure 4.1 Black Vs Non-black MDR analysis; one synergistic interaction was observed between rs8051733 and rs1129038 when adjusted for rs16891982

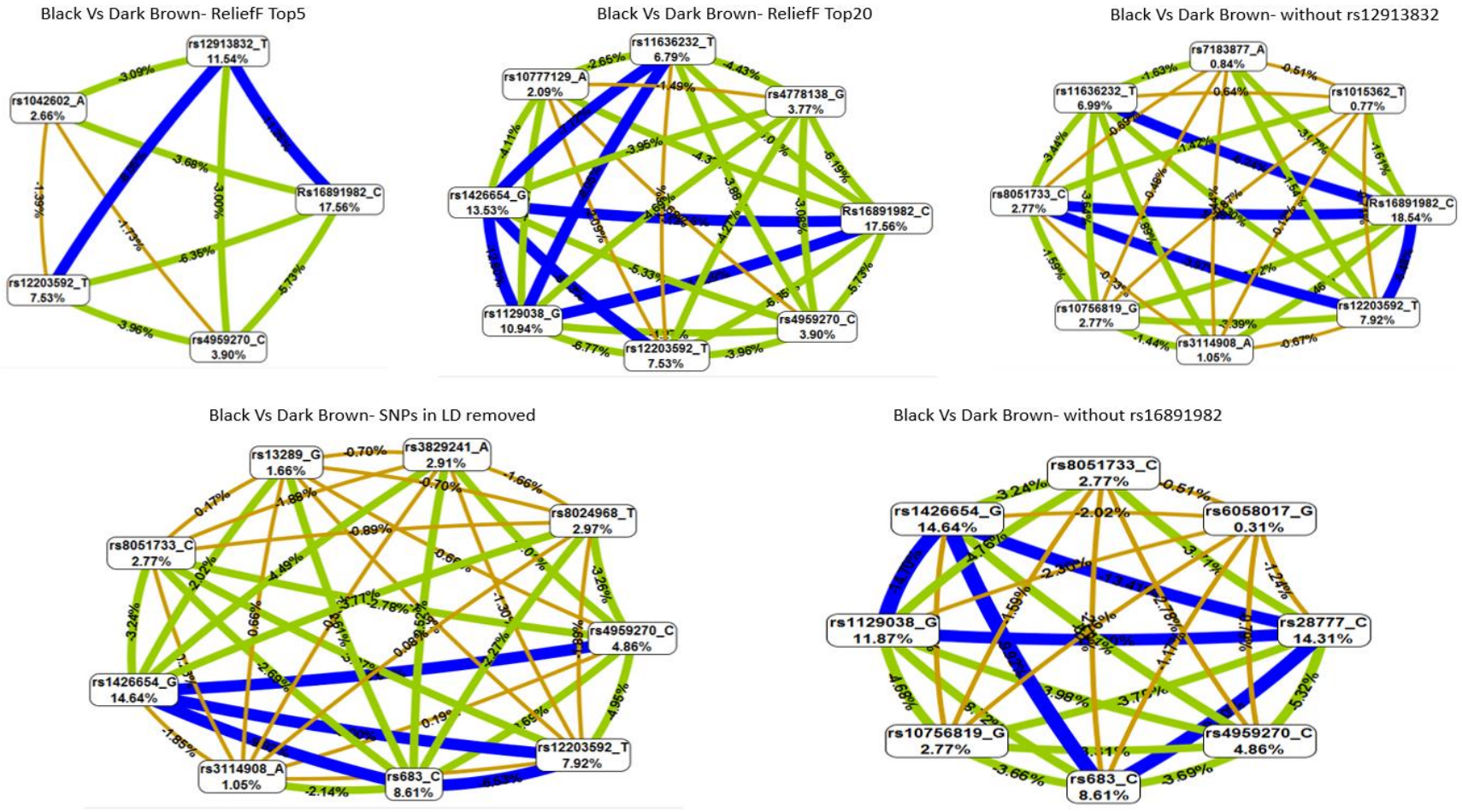

Figure 4.2 Black Vs Dark Brown MDR analysis; No synergistic interactions were observed

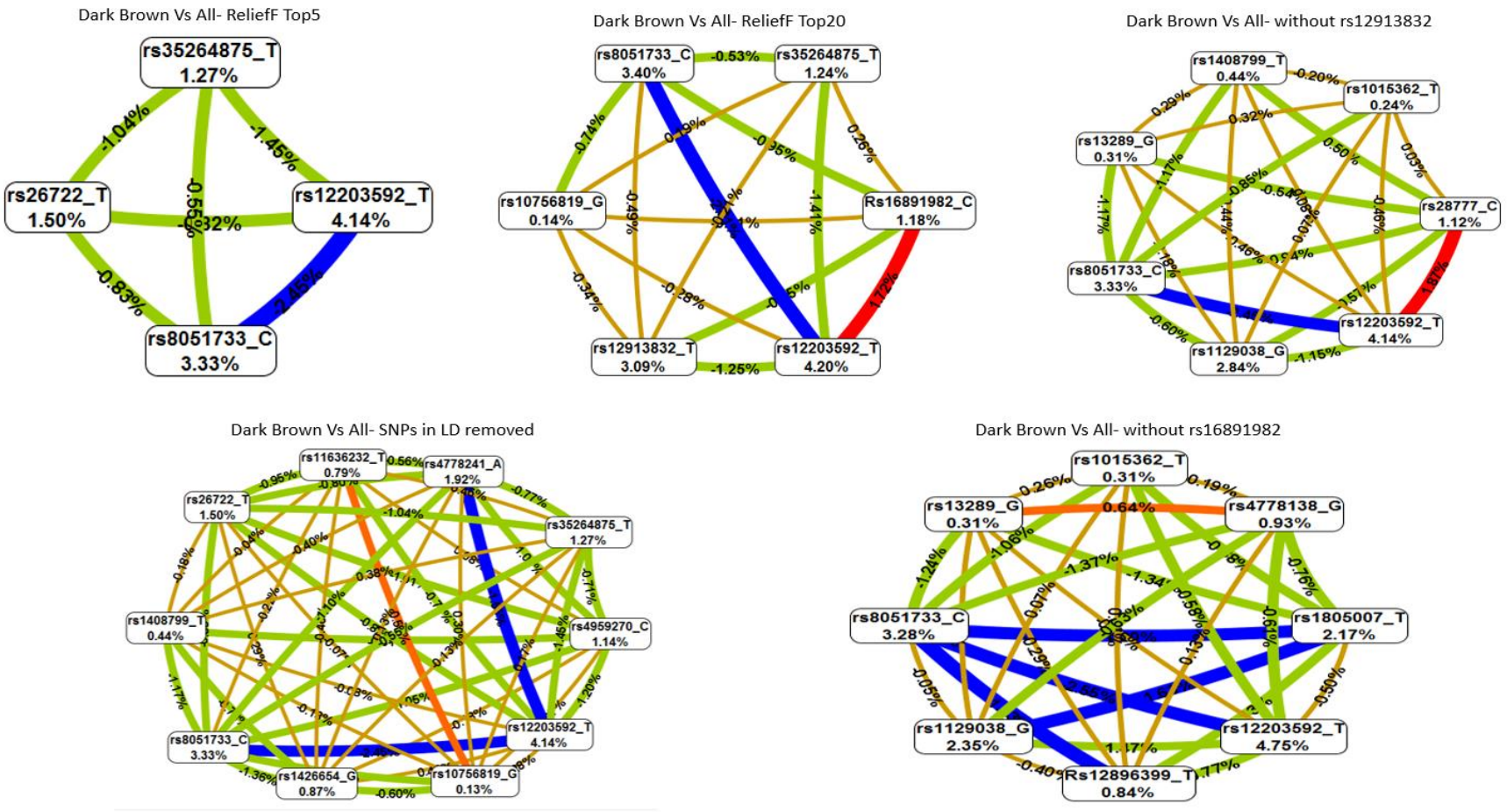

Figure 4.3 Dark brown Vs Non- Dark brown MDR analysis; Four synergistic interactions were observed between- rs12203592 and rs16891982 (using ReliefF 20 filter); rs12203592 and rs28777 (when adjusted for rs12913832), rs11636232 and rs10756819 (after removing all the SNPs in LD) and rs13289 and rs4778138 (when adjusted for rs16891982)

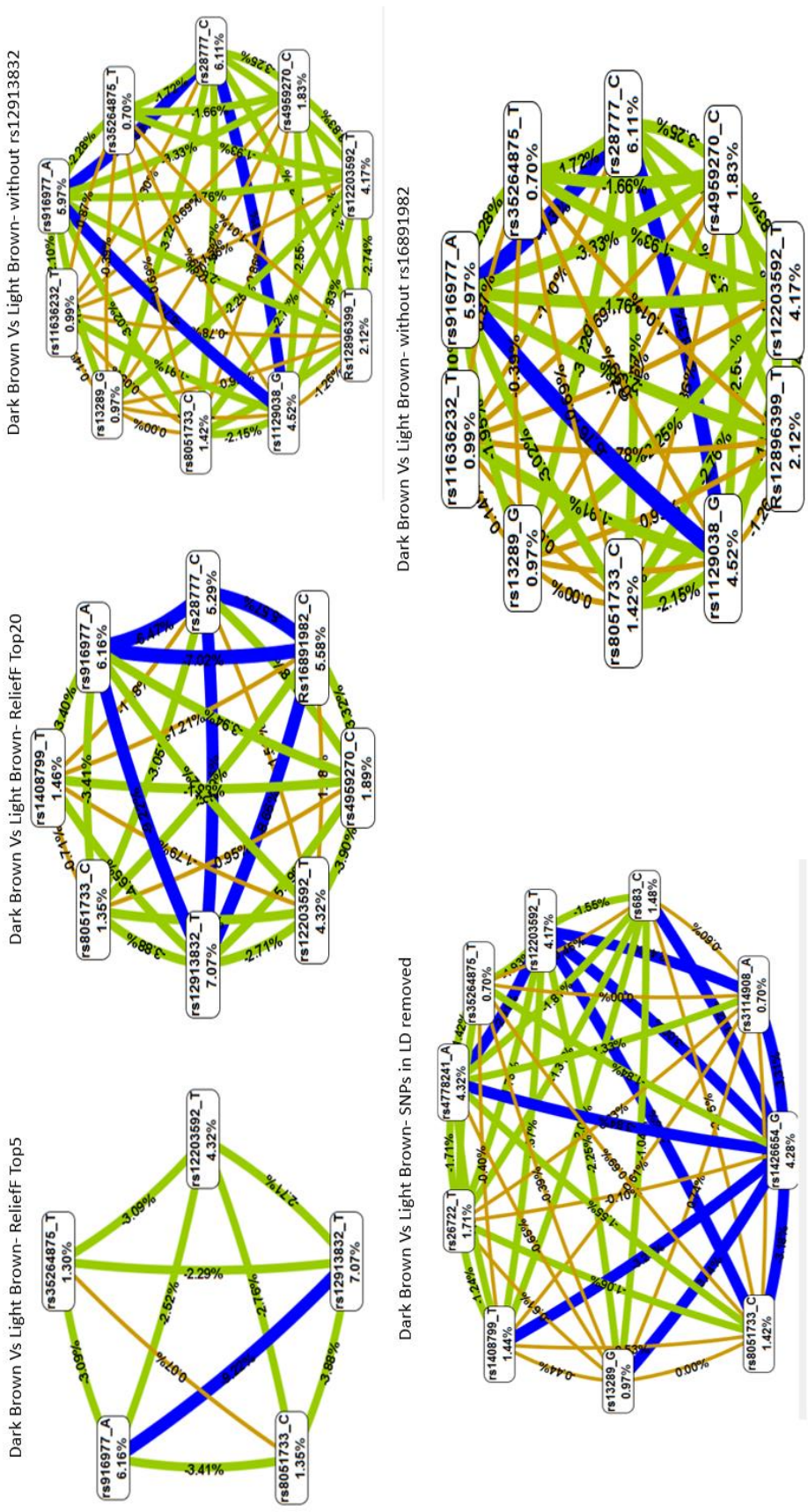

Figure 4.4 Dark brown Vs Light Brown MDR analysis; No synergistic interaction observed

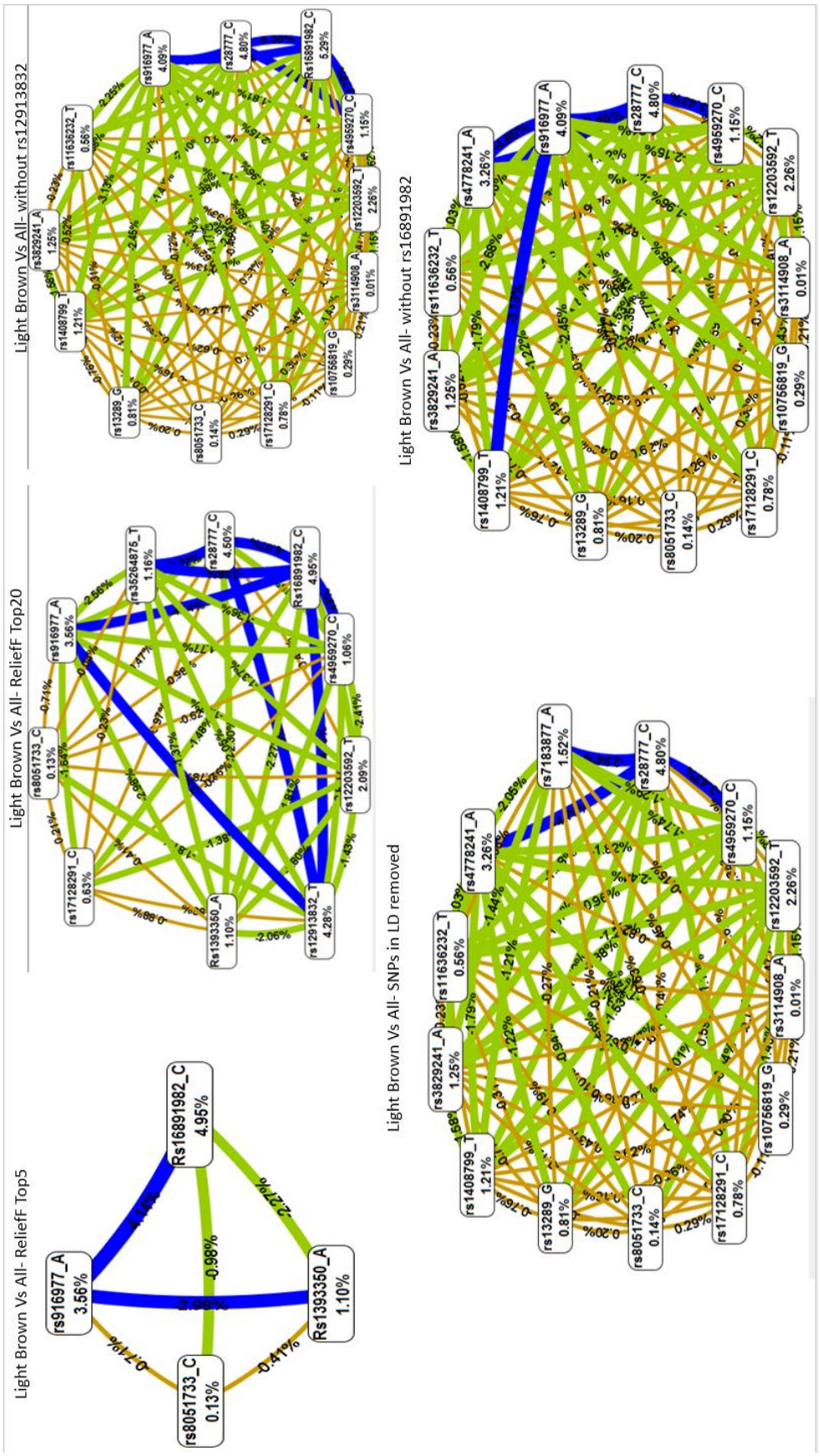

Figure 4.5 Light brown Vs Non- Light brown MDR analysis; No synergistic interaction observed

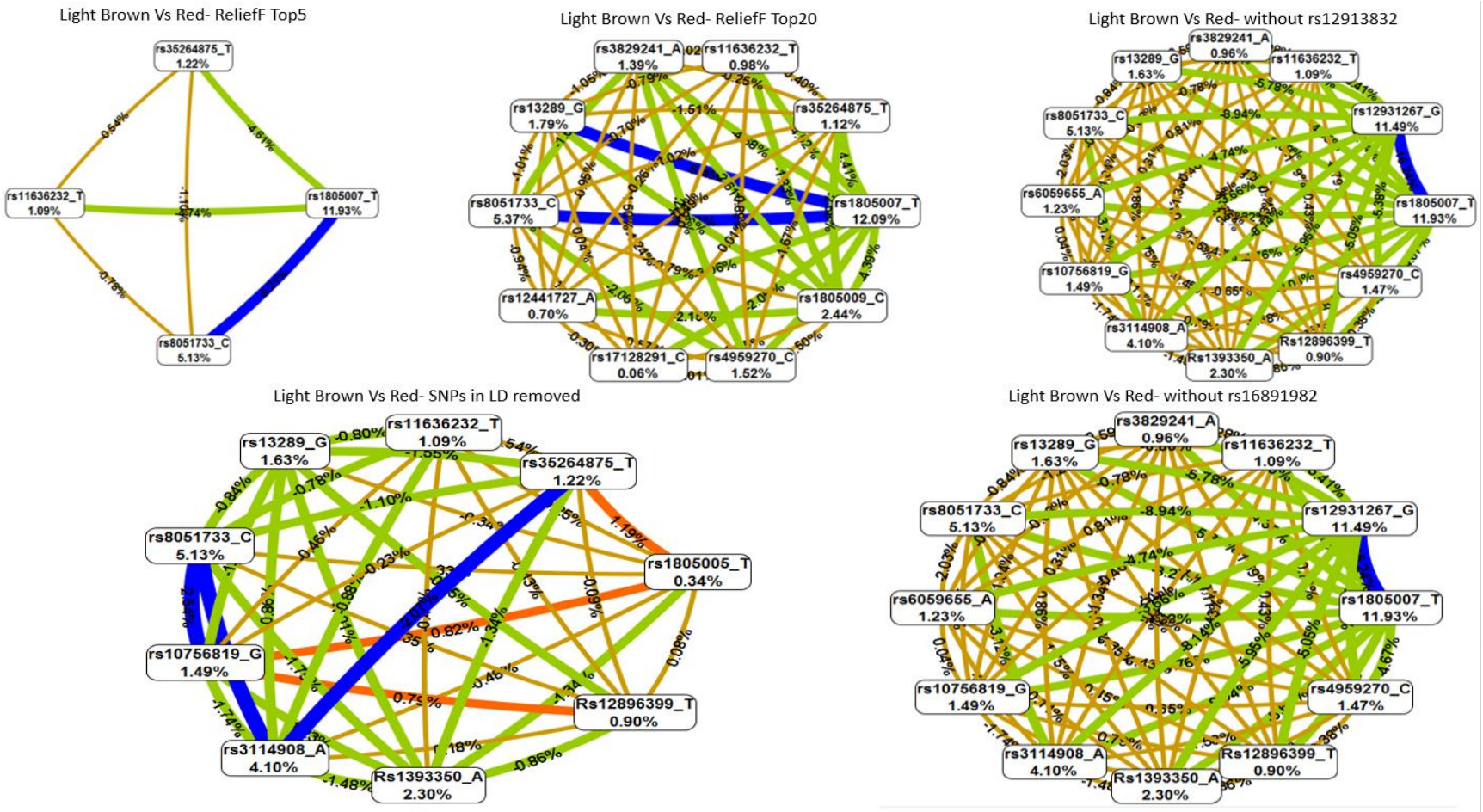

Figure 4.6 Light brown Vs Red MDR analysis; 3 synergistic interactions observed between rs35264875 and rs1805005, rs10756819 and rs1805005 and finally rs12896399 and rs10756819 after removing SNPs in LD.

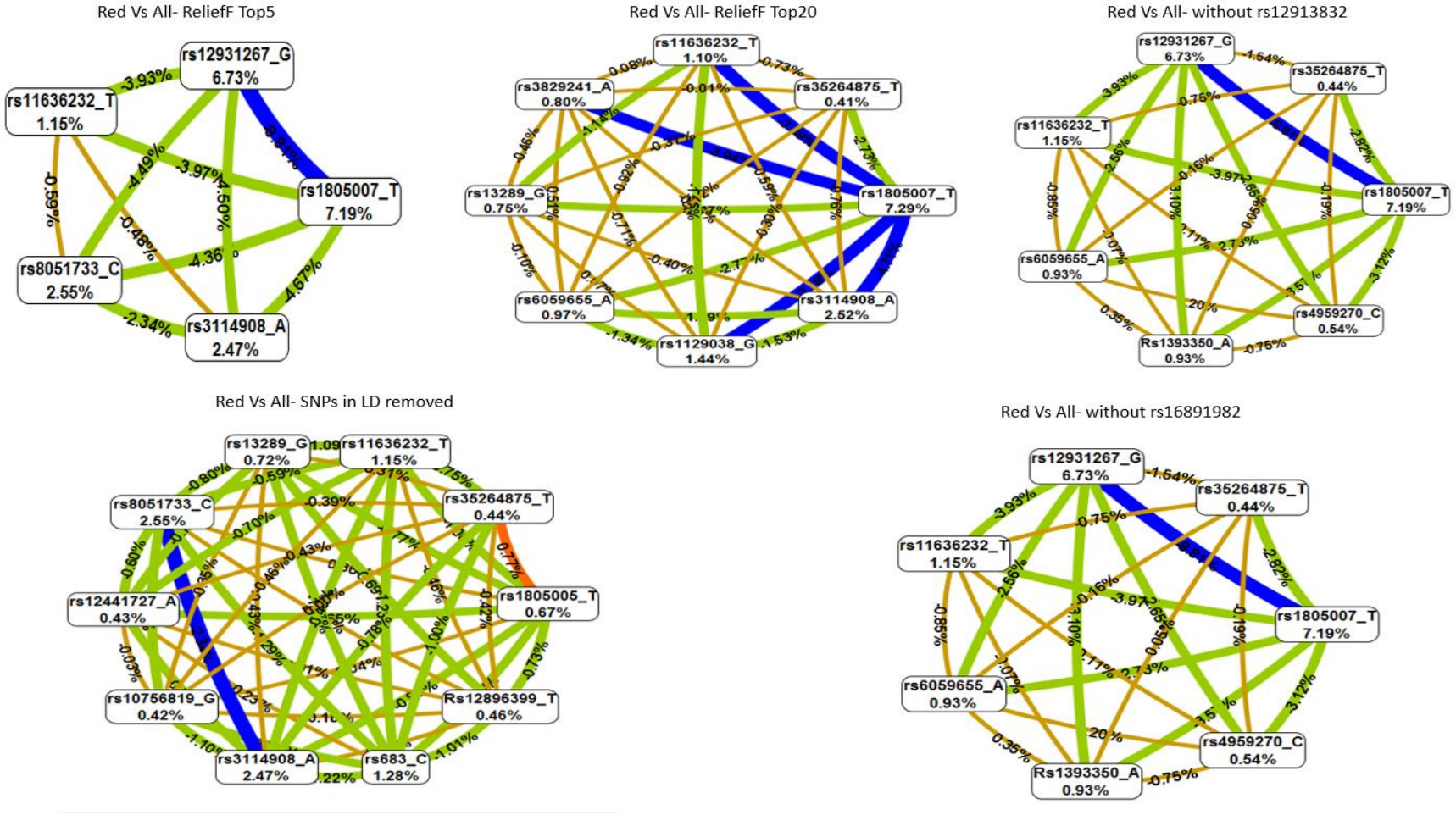

Figure 4.7 Red Vs Non- red MDR analysis; 1 synergistic interaction between rs35264875 and rs1805005 was observed after removing SNPs in LD. (Note: This interaction was also observed in Light brown Vs Red)

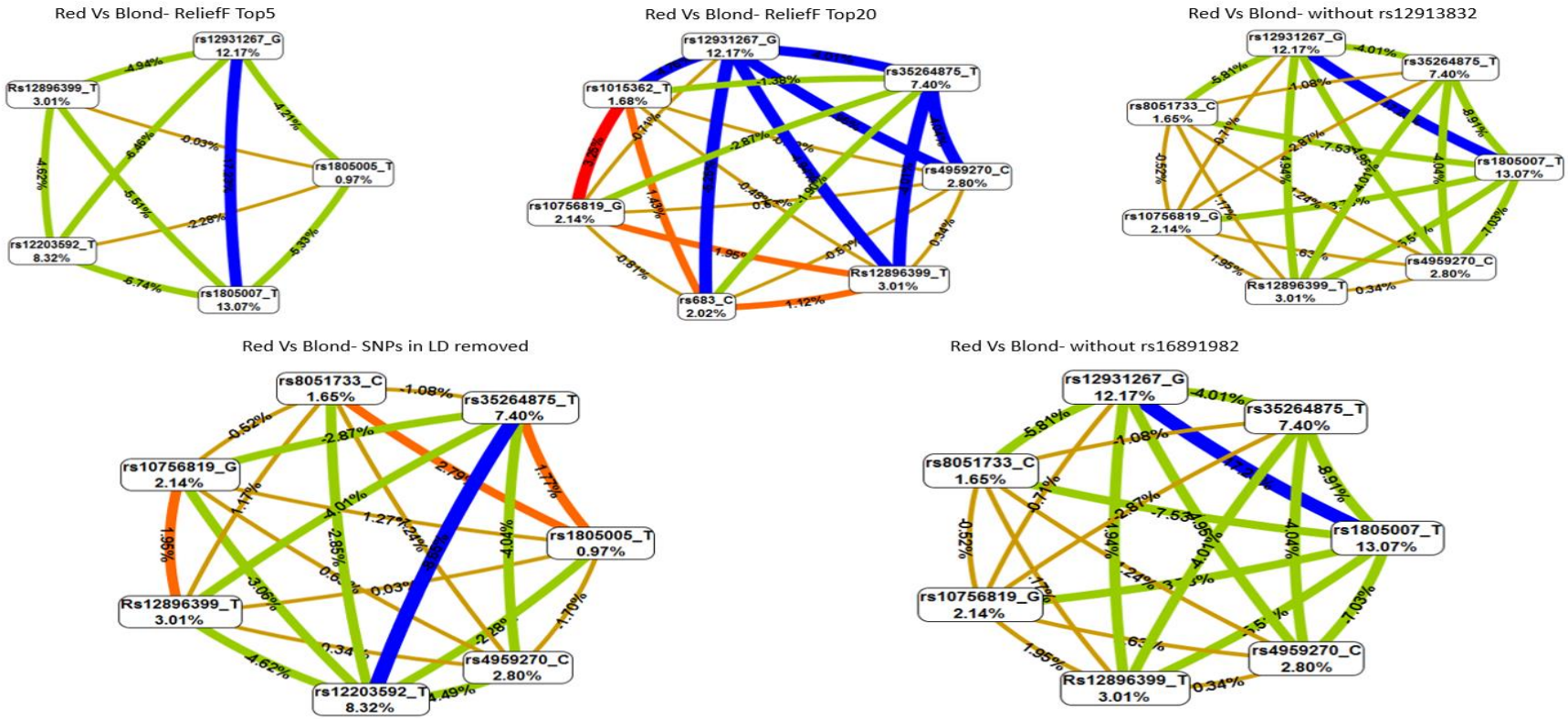

Figure 4.8 Red Vs Blond MDR analysis; four synergistic interactions were observed between- rs1015362 and rs10756819, rs1015632 and rs683, rs12896399 and rs683 and finally rs10756819 and rs12896399 when filtering top 20 SNPs using ReliefF filter. After removing the SNPs in LD, three interactions were synergistic between rs10756819 and rs12896399, rs8051733 and rs1805005 and finally rs35264875 and rs1805005 (Note: This interaction was also observed in light brown vs red and also red vs non- red).

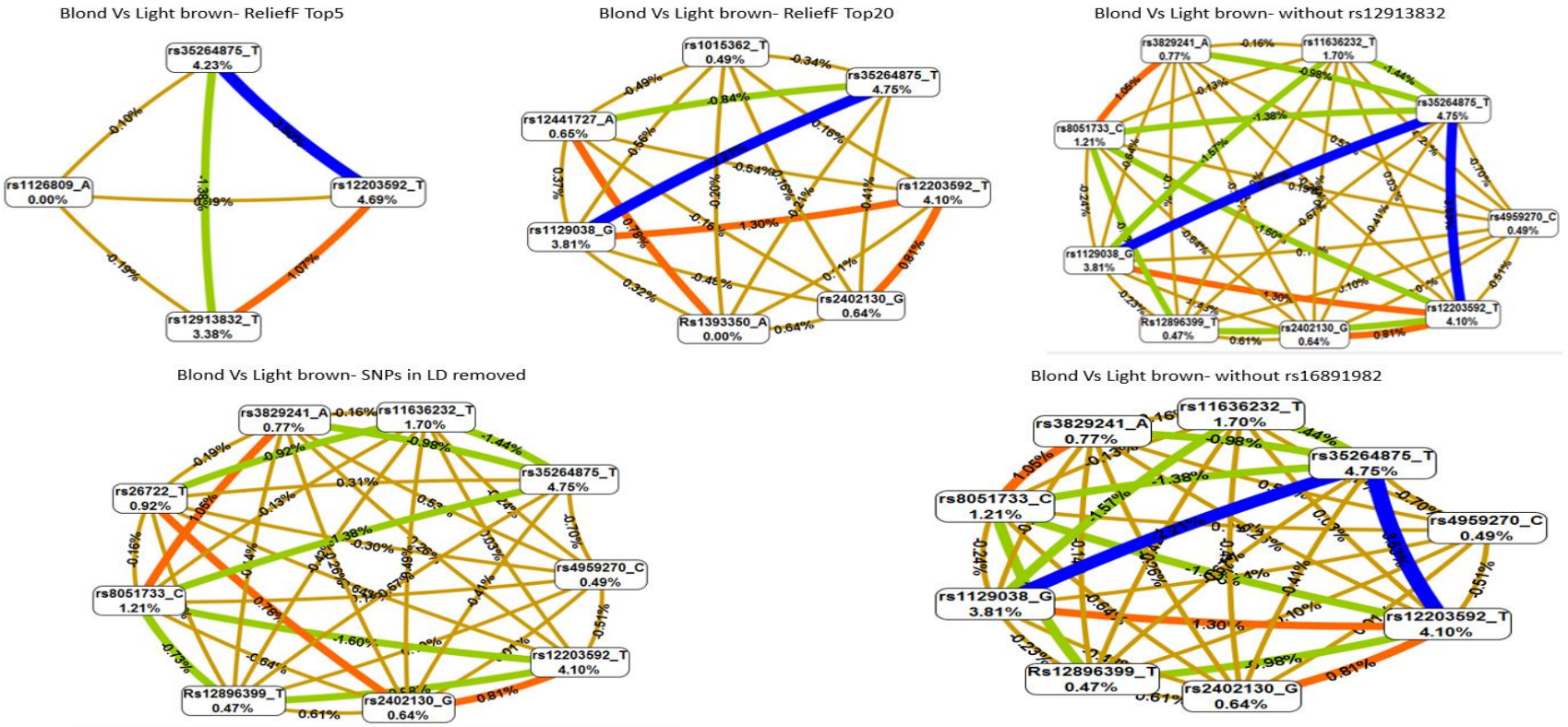

Figure 4.9 Blond Vs Light brown MDR analysis; about 13 interactions were synergistic. Interaction between rs12203592 and rs12913832 / rs1129038 (considered the same as they are in high LD) showed up under 3 different filtering methods, rs12203592 and rs2402130 showed up in 4 filtering methods, rs8051733 and rs3829241 showed up when adjusted for rs12913832 or rs16891982. The rest of the interactions were observed only once amongst 5 different analysis methods.

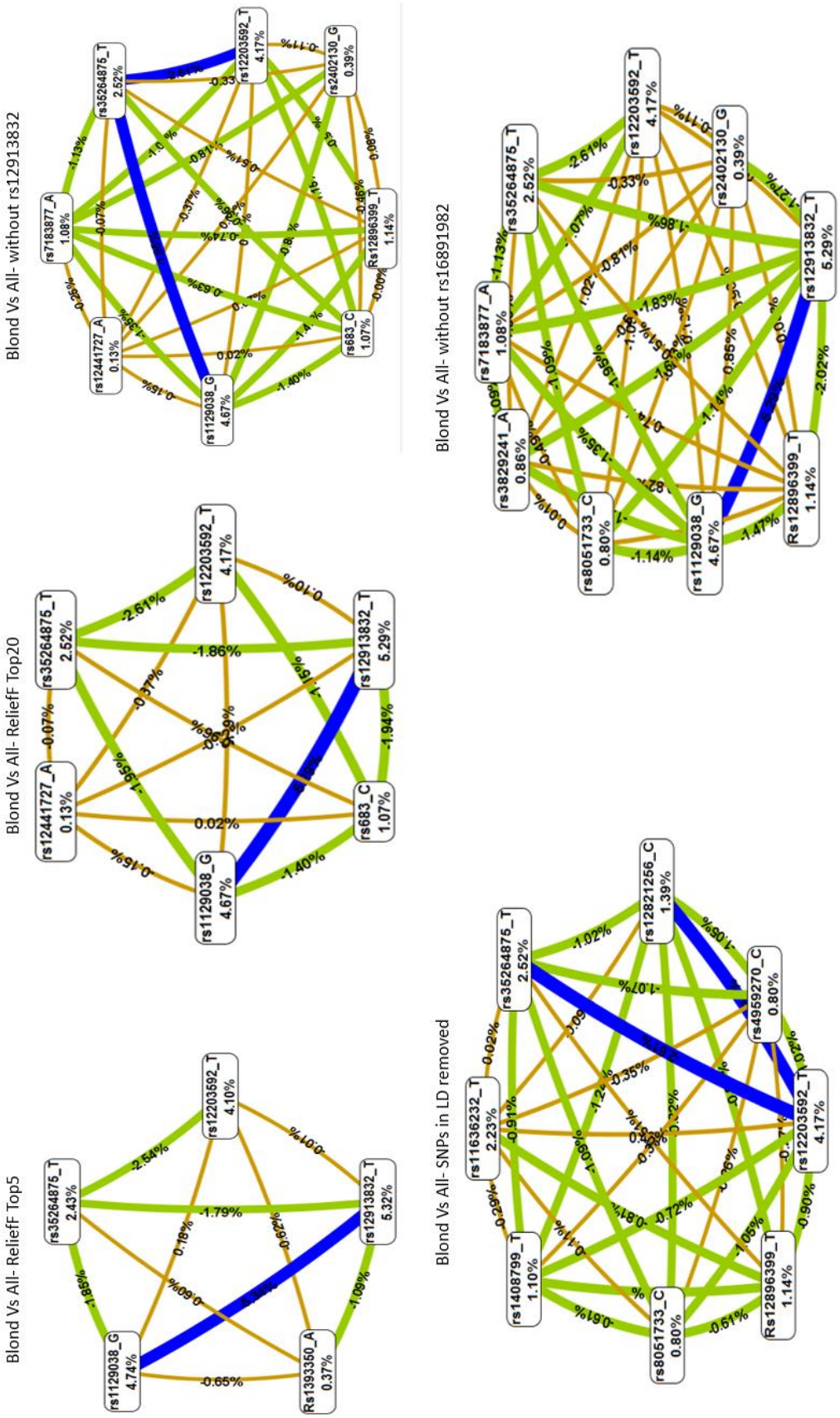

Figure 4.10 Blond Vs Non- blond MDR analysis; No synergistic interactions were observed.

Of all the MDR analyses, only 20 interactions were synergistic. So, a 1000-fold permutation testing was performed on each of these 20 SNP- SNP interactions in order to calculate the mean *p*-value. A table of all the synergistic interactions, trait classification and their corresponding *p*-values are listed in the table below (Table 4.1).

Table 4.1 MDR analysis for pairwise interactions for different trait classifications. Balanced accuracy refers to (senstitivity+specificity)/2; cross validation consistency refers to the number of times the model was observed when the data was split 10 times; * indicates *p*-values that were significant at α= 0.05 and number of permutations= 1000

Trait classification	**SNP1**	**SNP2**	**MDR- filter**	**Testing balance accuracy**	**CV consistency**	***P*-value**
Black/ Non-black	rs8051733	rs1129038	Without rs16891982	0.6562	10	0.002*
Blond/ Light brown	rs12203592	rs12913832	ReliefF- Top SNPs	0.6034	10	0.007*
	rs12441727	rs1393350	ReliefF- Top SNPs	0.6023	10	0.012*
	rs12203592	rs1129038	ReliefF- Top SNPs	0.5955	10	0.024*
	rs12203592	rs2402130	ReliefF- Top SNPs	0.5821	10	0.056
	rs8051733	rs3829241	SNPs in LD removed, without rs12913832 or rs16891982	0.6056, 0.6056	10, 10	0.013*, 0.012*
	rs26722	rs2402130	SNPs in LD removed	0.5688	8	0.174
Dark brown/ Non- Dark brown	rs12203592	rs16891982	ReliefF- Top SNPs	0.5524	10	0.061
	rs12203592	rs28777	Without rs12913832	<0.549	NA	>0.1

Table 4.1 continued

Dark brown/ Non- Dark brown	rs11636232	rs10756819	SNPs in LD removed	0.5653	10	0.026*
	rs13289	rs4778138	Without rs16891982	0.5569	9	0.046*
Red/ Non- red	rs35264875	rs1805005	SNPs in LD removed	0.5445	6	0.32
Red/ Blond	rs1015362	rs10756819	ReliefF- Top SNPs	0.6238	10	0.017*
	rs12896399	rs10756819	ReliefF- Top SNPs, SNPs in LD removed	0.6284, 0.6567	10	0.013*, 0.001*
	rs12896399	rs683	ReliefF- Top SNPs	0.6234	10	0.019*
	rs8051733	rs1805005	SNPs in LD removed	0.5922	9	0.129
	rs35264875	rs1805005	SNPs in LD removed	0.6307	10	0.018*
Red/ Light brown	rs10756819	rs1805005	SNPs in LD removed	0.5983	9	0.041*
	rs12896399	rs10756819	SNPs in LD removed	0.6499	10	0.001*
	rs35264875	rs1805005	SNPs in LD removed	0.6007	9	0.046*

Some of the interactions were redundant between different trait classifications. For example, rs35264875 (*TPCN2*) X rs1805005 (*MC1R*) and rs12896399 (*SLC24A4*) X rs10756819 (*BNC2*) had synergistic interactions with red/blond and also with red/light brown comparisons. 500 models were tested by MDR and correction for multiple testing was performed using Bonferroni correction. None of the interactions showed up as significant when Bonferroni correction was applied. As noted earlier, MDR-pt calculates *p*-values based on permutations. A study that compared multiple testing correction methods identified that permutation tests tend to identify more features than Bonferroni while still controlling for type- I errors[89]. Based on these permuted *p*-values, 12 interactions were found to be significant.

4.2 Binary Logistic Regression

As with MDR analysis, 10 different trait classifications were used for PLINK epistasis analysis in order to maximize the information. Nine interactions were found to be significant across different traits. Some of the interactions overlapped with MDR analysis. Depending on the size of the data set, the number of tests performed by PLINK ranged from 1398 to 1663. Bonferroni correction was applied for each analysis correcting for multiple testing. However, only two interactions were found to be significant after Bonferroni correction (see Table 4.2). Since, PLINK currently does not support permutation testing for SNP X SNP epistasis, Bonferroni correction was used for binary logistic analysis. PLINK epistasis also provides odds ratio for each interaction. It follows the standard convention where a value >1 is considered favorable towards the 'cases'/ success, a value < 1 is considered to favor the 'controls' /failures, and a value of 1 is considered to have no or equal effect on cases and controls and therefore is not useful[90]. As there were many SNP- SNP interactions that were significant at $\alpha = 0.05$ for each analysis, only interactions that were significant after Bonferroni correction are shown in Table 4.2. All the significant interactions observed via PLINK are listed in Appendix B.

Table 4.2 PLINK analysis for different trait classifications. Only interactions with significant *p*-value after correction for multiple testing are reported in the table.

Trait Classification	**SNP1**	**SNP2**	**Odds Ratio of Interaction**	**PLINK (Bonferroni corrected)**
Black/ Non- black Black/ Dark brown	rs12203592 *(IRF4)*	rs4959270 *(EXOC2)*	6.924 0.1572	0.000006581 2.36E-05
Black/ Non- black	rs12203592 *(IRF4)*	rs35264875 *(TPCN2)*	3.703	0.00001499

In order to ensure that the SNPs in significant interactions are not in LD, PLINK epistasis was performed again by pruning SNPs in LD. LD based pruning of the dataset was performed after identifying SNPs in LD via a sliding window procedure. The threshold was set at $r^2 = 0.75$ and, for each SNP pair in which LD was > 0.75, one of the SNPs was pruned out. For example, in black vs dark brown case-control analysis, 6 SNP- SNP pairs were in high LD at $r^2 > 0.75$. One such combinations was rs12913832 (*HERC2*) and rs1129038 (*HERC2*) at $r^2 = 0.9124$, which is a well-known to be in high LD [91]. Hence, one of the SNPs (rs1129038) was removed for epistasis analysis after LD pruning. The results largely remained the same even after LD pruning.

Summary of MDR and PLINK analysis:

In summary, 23 interactions were identified for hair color. Of which, 2 were significant after Bonferroni correction. A final summary of all the interactions obtained from MDR and PLINK analysis is given in Appendix C.

4.3 Predictive Modelling

1000 cross-validations (CV) with randomized training and test sets were performed and the performance was assessed by taking the average of the 1000 CV AUCs generated. SNPs and interactions were then ranked based on 'all samples' AUC calculations (An improvement of at least 0.5% in AUC in any category or 1% increase in overall AUC in comparison to HIrisPlex was looked for). Next step was to check the significance of co-efficients. If the SNP or interaction had co-efficients that had *p*-value < 0.05 in any of the categories, they were retained. Then, the SNPs and interactions that decreased the AUC (in comparison to cross-validated HIrisPlex model) after 1000 cross validations were rejected. Finally, a list of SNPs or interactions that could be modeled into the HIrisPlex system was made. The table containing the AUC values from 'all samples AUC runs' and 1000 cross-validation runs are given in Appendix D.

A Pearson correlation matrix was also generated for all the interaction diplotypes and also for the SNPs involved in the interaction in order to assess their correlation to individual hair colors using SPSS v. 24 (IBM Corp. New York, USA) (Appendix E).

After assessing the AUC values post cross-validation, 5 SNPs or interactions were found to increase the accuracy of prediction. These were ranked based on the percentage of AUC increase (Table 4.3).

Table 4.3 List of models containing SNPs/ Interactions that add on value to HIrisPlex model post cross-validation. # indicates- HIrisPlex+ An interaction observed within HIrisPlex SNPs; * indicates- HIrisPlex+ 1 SNP; ** indicates- HIrisPlex+ SNP1 + SNP1's interaction with one of HIrisPlex SNPs; *** indicates- HIrisPlex+ SNP1+ SNP2+ Interaction between SNP1 and SNP2

SNPs/ Interactions that increase AUCs by at least 0.5% for individual categories and by at least 1% for overall AUC in 100% AUC (all samples) calculation	**Ranking based on 100% AUC**
HirisPlex+rs35264875+rs12203592*rs35264875	1**
HIrisPlex+rs26722	2*
HirisPlex+rs12896399*rs683	3#
HirisPlex+rs1015362+rs10756819+rs1015362*rs10756819	4***
HirisPlex+rs12441727+rs12441727*rs1393350	5**

In order to avoid over fitting, each of these SNPs or interactions was added sequentially to the HIrisPlex model to assess performance. Since the SNPs found in interaction, rs12896399*rs683 (*SLC24A4*TYRP1*), are already in the HIrisPlex model, the multiplied diplotype of this interaction was added first although it was ranked 4th on accuracy increase. Each model was added on top of the existing model and then AUCs and cross-validated AUCs were calculated. For example, rs35264875 (*TPCN2*) and its interaction, rs12203592*rs35264875 (*IRF4*TPCN2*) were added on to the first model which consisted of HIrisPlex+ rs12896399*rs683 (*SLC24A4*TYRP1*) (HIrisPlex SNP-SNP interaction).

Increase or decrease in AUC value was assessed for each new addition of SNP-SNP interaction by comparing the values to the previous model and also to HIrisPlex.

The SNPs or interactions that decreased the AUC of the previous model after cross-validation were dropped. For example, the addition of rs12441727 (*OCA2*) and its interaction-*OCA2* rs12441727 * *TYR* rs1393350 decreased the AUC value for black and blond categories after cross-validation, hence this combination was rejected. Table 4.4 Lists the AUCs for each sequential addition. A table with all model values is provided in Appendix F.

Table 4.4 AUC values (before and after cross-validation) for sequential addition of SNPs and interactions to HIrisPlex model. The model highlighted in yellow was found to increase AUC across all categories especially with red hair color.

Models	Model #	Black100	Black90	Brown100	Brown90	Red100	Red90	Blond100	Blond90
HirisPlex	1	0.9473	0.8625	0.8516	0.7616	0.9575	0.9121	0.9012	0.8454
HIrisPlex+rs12896399X683+rs35264875_T+rs12203592X35264875	2	0.9508	0.8696	0.8698	0.7728	0.9648	0.9116	0.9160	0.8501
HIrisPlex++rs12896399X683+rs35264875_T+rs12203592X35264875 +rs26722_T	3	0.9583	0.8748	0.8770	0.7769	0.9653	0.9147	0.9159	0.8485
HirisPlex+rs12896399X683+rs35264875_T+rs12203592X35264875+ rs26722_T+rs1015362_T+rs10756819_G+rs1015362X10756819	4	0.9618	0.8693	0.8878	0.7780	0.9725	0.9209	0.9219	0.8461
HirisPlex+rs12896399X683+rs35264875_T+rs12203592X35264875+ rs26722_T+rs1015362_T+rs10756819_G+rs1015362X10756819+rs1 2441727_A+rs12441727X1393350	5	0.9652	0.8430	0.8916	0.7662	0.9759	0.9208	0.9217	0.8366

The final model chosen increased AUC values across all four categories when compared to HIrisPlex after adding all informative SNPs and interactions (see Figure 4.12 and Table 4.5). To estimate the performance and accuracy of the new model in comparison to the existing HIrisPlex model, Receiver Operating Characteristics (ROC) curve was used. ROC curves were plotted in RStudio. The test consisted of evaluating the models using AUC (Area Under the ROC Curve) values with the inclusion of all the samples in the dataset. AUC values range from 0.5 to 1, where 0.5 indicates random prediction and 1 indicates accurate/perfect prediction[92].

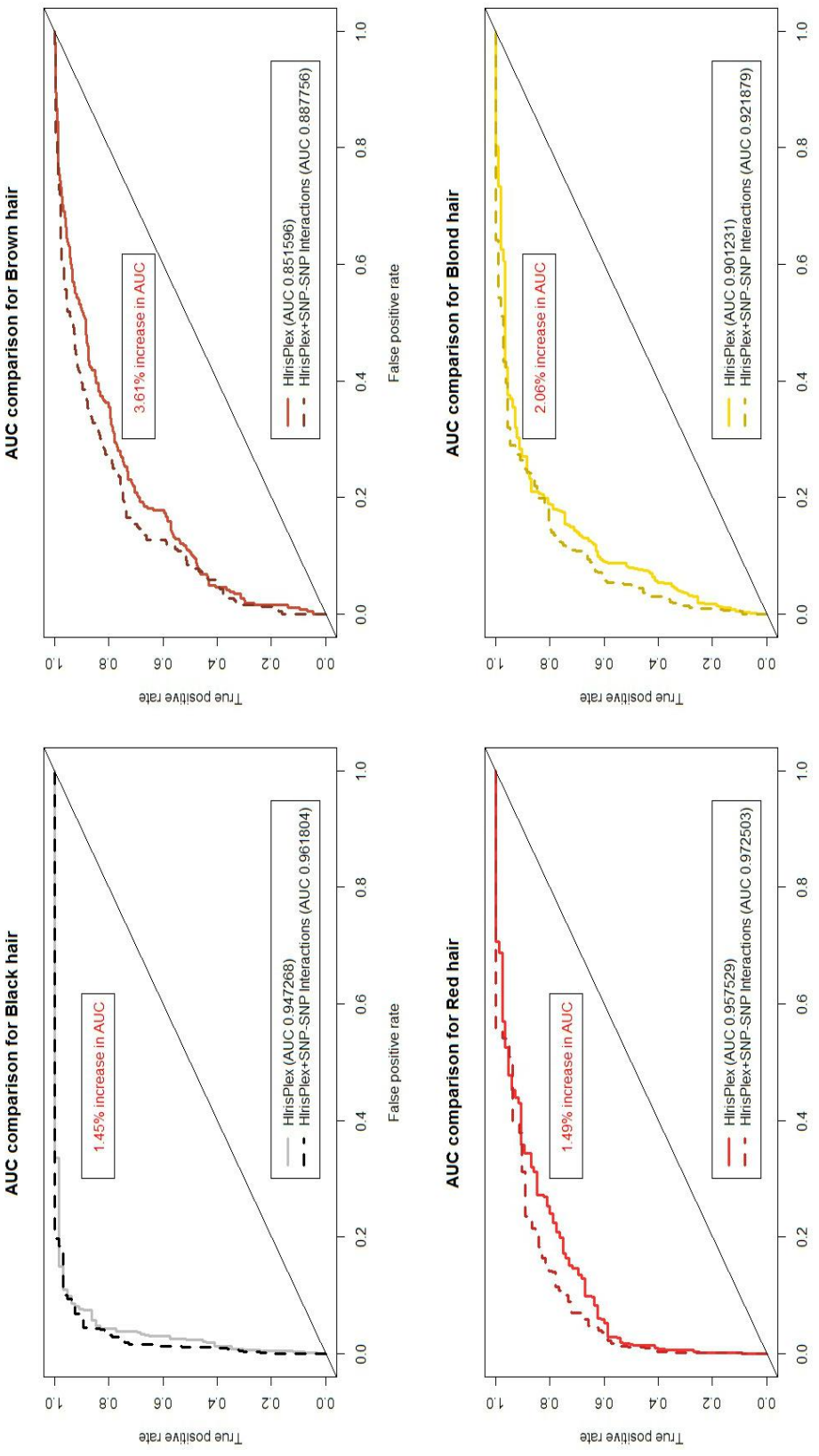

Figure 4.11 ROC curve of hair color prediction models using 100% training and test set. The increase in AUC by including SNP-SNP interactions into the HIrisPlex model is noted in red.

The final model includes the following SNPs and interactions- *SLC24A4* rs12896399 x *TYRP1* rs683, *IRF4* rs12203592 x *TPCN2* rs35264875, *ASIP* rs1015362 x *BNC2* rs10756819 and rs26722 (*SLC45A2*). In order to account for the independent effects, the SNPs in each interaction were also added to the overall model. An overview of its performance can also be found in Table 4.5.

Apart from the interactions, rs26722 (*SLC45A2*) had an independent effect on hair color in this dataset. A Friedman rank sum test indicated the differences in AUC between HIriPlex and the new model was statistically significant (Friedman chi-squared = 4, df = 1 , p-value = 0.0455). Distributions of three identified interactions and one independent S NP with hair color categories are listed in Appendix G.

Table 4.5 Hair color prediction accuracies displaying AUC using 100% dataset, and 10% dataset (mean of 1000 cross validations) in HIrisPlex SNPs alone versus a combination of HP with the addition of SNP-SNP interactions. Improvements in AUC for the independent test are displayed in red. * This also included the independent effect found for rs26722 (*SLC45A2*) in this dataset.

		HIrisPlex model		**HIrisPlex including significant interactions***	
		100% dataset	10%, mean AUC (CV 1000)	100% dataset	10%, mean AUC (CV 1000)
Hair Color	**Black**	0.947	0.863	0.962	**0.869**
	Brown	0.852	0.762	0.888	**0.778**
	Red	0.958	0.912	0.973	**0.921**
	Blond	0.901	0.845	0.922	**0.846**

Of the SNPs in the SNP-SNP combinations, rs1015362 (*ASIP*) and rs26722 (*SLC45A2*) are included in Snipper suite 2.5's eye color prediction tool, rs35264875 (*TPCN2*) is included in their hair color prediction model, and rs10756819 (*BNC2*) is included in HIrisPlex- S (skin color prediction tool, paper in preparation). rs12203592 (*IRF4*), rs12896399 (*SLC24A4*) & rs683 (*TYRP1*) are already in the HIrisPlex system [5-8].

Hair color and pigmentation in general are complex polygenic traits. Many GWAS studies have identified markers that contribute an independent effect on pigmentation prediction. It is interesting to note that SNPs identified as informative markers for other pigmentation prediction systems, seem to have an interactive effect with the hair color informative SNPs, thereby increasing the statistical accuracy of prediction. Since there is no 'recommended' analysis method for epistatic analysis, two different approaches were considered. In this study, 61 pigmentation markers were evaluated using MDR and BLR methods and three epistatic effects (Figure 4.12) were identified between *SLC24A4* rs12896399 x *TYRP1* rs683, *IRF4* rs12203592 x *TPCN2* rs35264875, *and ASIP* rs1015362 x *BNC2* rs10756819.

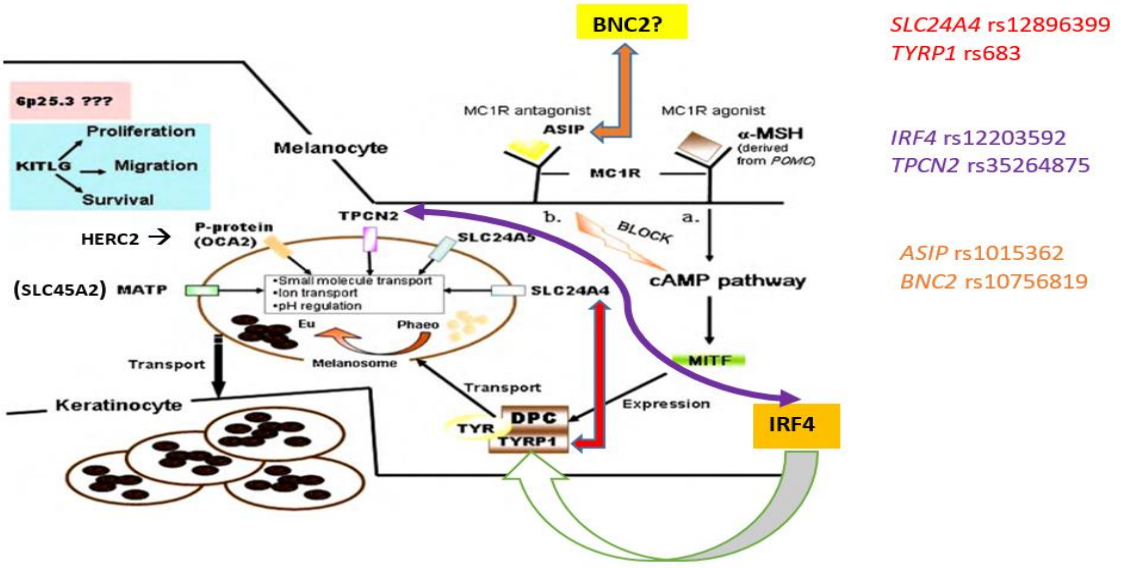

Figure 4.12 Pigmentation pathway with proposed interactions; Proposed interactions between the genes are represented in orange, red and purple arrows. Image adapted from Scherer & Kumar, 2010

SLC24A4 (Chromosome 14) is a gene that encodes for proteins located in the membrane of melanosome and regulate ionic concentration/ pH whereas, *TYRP1* (Chromosome 9) is one of the key enzymes that catalyzes the synthesis of melanin, specifically eumelanin within the melanosomes [93]. Eumelanins are synthesized when there is a drop in cysteine concentration. *OCA2* is a well-known regulator of pH and the presence of a mutation or absence of this gene leads to impaired flow of Cl^- ions thereby leading to lighter pigmentation. Perhaps, *SLC24A4* (ion channel) functions in a way that triggers the eumelanin pathway to counteract the effect of ion imbalance introduced by *OCA2.* SNP-rs12896399 (*SLC24A4*) is in the intronic region and its interacting SNP –rs683 (*TYRP1*) is in the exonic (3' UTR) region. Perhaps, this is another case of a SNP in an intronic region having an effect in the regulation of expression of another gene. Although the genes are located in different chromosomes, similar to an observation made by Steiner *et al.,* on *MC1R* and *Agouti* on mouse models[94], it could be possible that depending on the allele in rs12896399 (gene), it could up/ down regulate *TYRP1* leading to corresponding darker/ lighter hair color.

In the second interaction, *IRF4* rs12203592 x *TPCN2* rs35264875, rs12203592 is located in an intronic region, whereas rs35264875 is a missense mutation. Previously, the role of *IRF4* (Chromosome 6) in melanogenesis was unknown. However, recent studies [34, 35, 53] show that SNP rs12203592 has been associated with the brown hair phenotype, and reduced expression of the gene *IRF4* has been associated with lighter hair color in humans. In some species such as mice, *MITF* is also reduced when the hair color is lighter. It has been shown that rs12203592 regulates expression of IRF4, which interacts with MITF and in turn induces *TYR* (tyrosinase). *TPCN2* (Chromosome 11) has been associated with hair color previously (blond Vs brown hair) and the possibility of it having a role in coding for a putative ion channel releasing Ca^{2+} ions is still being investigated. If, it is confirmed that this gene also plays a part in ion transport [95], it is highly likely that this gene has an impact on melanogenesis via regulation of pH (similar to *OCA2* and *SLC45A2*). Perhaps, this intronic SNP regulates tyrosinase's functionality in the melanosome by

regulating the Ca^{2+} ion transporter, *TPCN2.* Since pH of the melanosomes decide whether pheo/eumelanin is being synthesized which then equates to the phenotype of light/ dark hair, it is likely that TPCN2's functionality is probably altered by IRF4's effect, therefore causing a change in the pH.

With reference to the third interaction, *ASIP* rs1015362 x *BNC2* rs10756819, the SNP in *ASIP* (Chromosome 20) is intergenic and the SNP in *BNC2* (Chromosome 9) is intronic. *ASIP* is a gene, coding for agouti signaling protein, which acts as an antagonist to melanocortin receptors [96]. *ASIP* binds to *MC1R* thereby decreasing the levels of cyclic adenosine monophosphate (cAMP) which is essential for the stimulation of *TYR* enzyme[97]. *BNC2* gene codes for Zinc finger protein basonuclin-2 and has been associated with pigmentation especially with continuous variation in European skin color[43]. Few researchers wanted to identify the biological significance of rs10756819 in *BNC2* and its functional relevance to pigmentation (skin color in specific) as this SNP is also located in an intronic region. They wanted to test if this intronic SNP was associated with differential gene expression of *BNC2*. However, they found that SNP rs12350739, which is in high LD with rs10756819, is the causal variant that affected the gene's expression. Apparently, *BNC2* is one of the most conserved genes expressed in many of the human tissues and could be a putative transcription factor [98]. Overexpression of *BNC2* in mice resulted in melanocytic cell death subsequently leading to pigment loss [98]. Variants in *BNC2* have been associated with freckling pattern [99]. Interestingly, the interaction that we observed, *ASIP* rs1015362 x *BNC2* rs10756819, improved red hair prediction accuracy the most. Also, the gene that has been mostly associated to red hair, *MC1R,* has also been reported to be associated to freckling. The similarity in phenotypes (freckling and red hair) between these genes suggest that *BNC2* needs to be studied further to understand its implications in producing the red hair phenotype and its functional interaction, specifically with *MC1R*. In fact, Praetorius *et al.,* suggested that *BNC2* could probably have an effect on *ASIP* which in turn triggers the pheomelanin pathway, thereby resulting in red hair phenotype [99]. Most of the independent markers that are associated with

pigmentation are intronic which regulates the expression of the gene (in some cases, a different gene) thereby leading to different phenotypes. The interaction between *HERC2* and *OCA2* is a classic example, where an intronic SNP rs12913832 in *HERC2* (depending on the type of allele found), led to binding of transcription factors such as *MITF* to the enhancer of *HERC2* which in turn elevated the expression of *OCA2* thereby causing a darker pigmentation [19]. The functional significance of these correlations/associations of epistatic interactions need to be further studied. Nevertheless, these bioinformatically-identified interactions seem to increase the accuracy of prediction in HIrisPlex prediction system across 4 hair color categories which looks very promising to research. Lastly, it is apparent from this study that there is a clear benefit to consider the inclusion of significant SNP-SNP interactions in future prediction model-building approaches.

CHAPTER 5. RESULTS & DISCUSSION: HAIR FOLLICLE PIGMENT GENE EXPRESSION

32 candidate genes were tested for differential gene expression patterns for this study. However, out of the 32 primer sets, only 26 (listed in Table 5.1) were amplifiable in human hairs even though their primer design catered for humans and owl monkey primates.

Table 5.1 Panel of genes that were assessed using qPCR

Genes that were used in qPCR analysis			
PPFIBP1	PLXNB1	IRF2	SLC1A2
C10orf11	SLC30A1	IRF2BP2	SLC16A9
PLXNC1	TMEM50B	TYRP1	SLC25A23
TRPM	KITLG	MIR29B2	KIT
C19orf44	MLANA	TYR	SLC16A1
PMEL	SLC7A5	SLC2A8	SLC34A2
SLIT3	SNX1		

Of these 26 genes, three of the genes; *SLC34A2, SLIT3, and SNX1* did not amplify for certain samples after repeated attempts. Hence, these were not included in further analysis. TBP was used as the housekeeping gene (HKG) for analyses and was multiplexed with every gene. Average C_T was calculated for HKG for each of the hair colors and the genes of interest were normalized using the corresponding average HKG C_T value. E.g. the average for HKG for blond hair color was 28.9565, dark brown was 29.8665, and red was 29.97939 (Figure 5.1). Depending on the hair color sample used, the gene of interest was normalized using the corresponding average HKG.

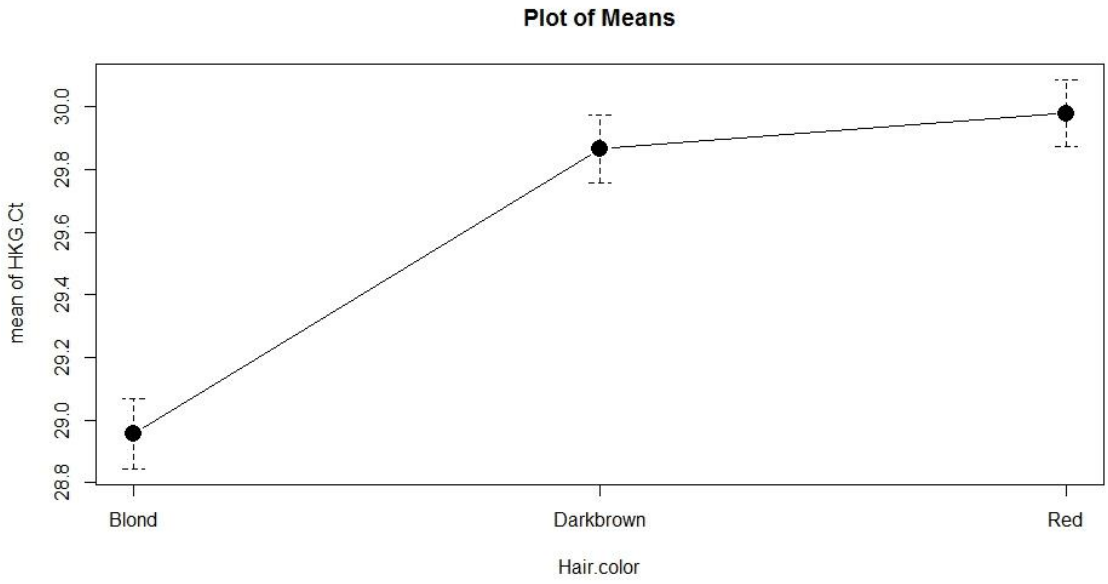

Figure 5.1 Plot of means of CT of housekeeping genes with standard error bars across different hair colors

Relative fold expression changes were calculated by log transforming $\Delta\Delta C_{T.}$ The fold change was assessed for statistical significance for each hair color in comparison to dark brown (Figures 5.2).

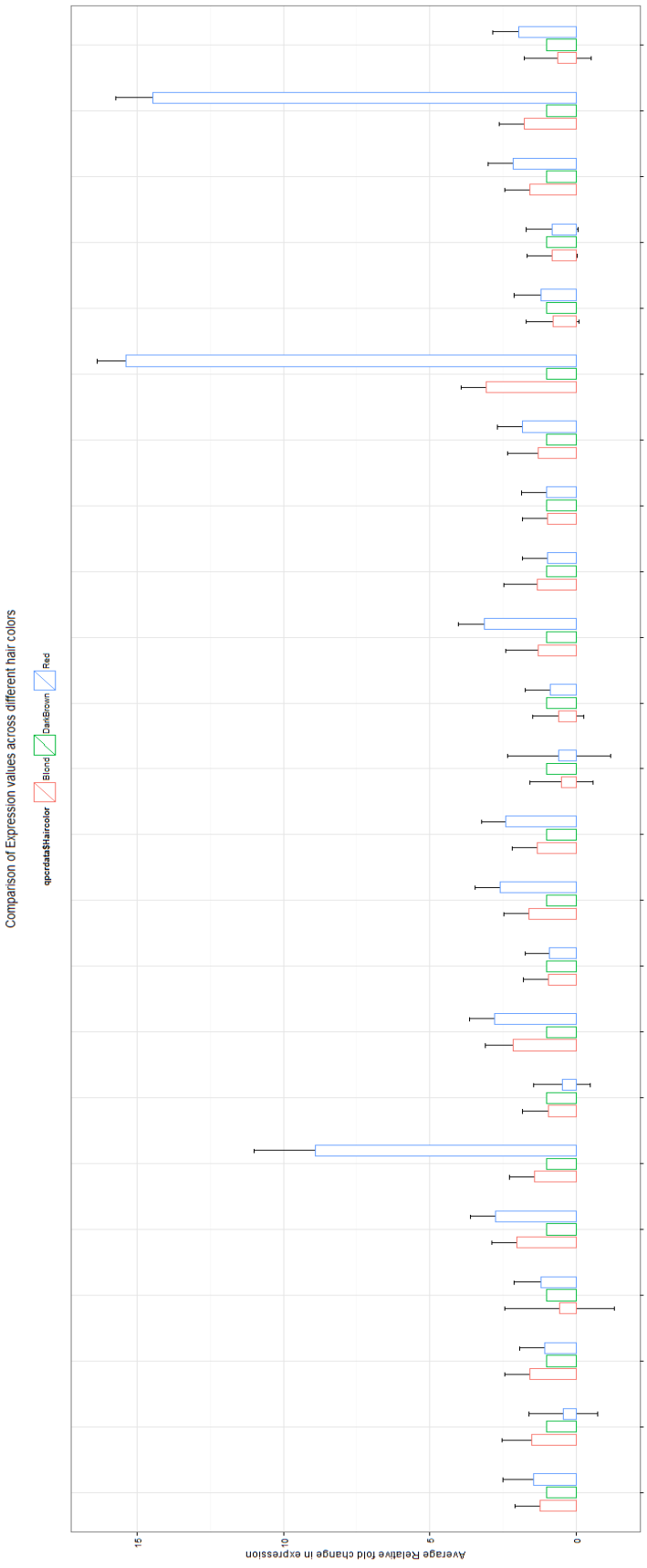

Figure 5.2 Relative fold change comparison. Comparison of relative fold change in expression of red and blond hair samples with respect to dark brown hair samples. (Bars indicate standard error). Plot created using RStudio.

Of the genes that amplified, *IRF2BP2* was found to be statistically significant with differing gene expression across hair color categories with a *p*-value= 0.0377. *IRF2BP2* is an Interferon regulatory factor 2 binding protein 2 located on chromosome 1 which interacts with *IRF2*. The expression of this gene in blond hair was less than dark brown whereas its expression was slightly higher in red. *IRF2BP2* was identified as a regulator of *NFAT1's* (Nuclear Factor of Activated T cells) transcriptional activity [100] where *NFAT1* has been identified to regulate hair growth and hair cycle. In fact, its family protein- *NFAT2* has been seen to be up-regulated while *MITF* and *TYR* expression were decreased in hypo-pigmented B16 melanoma cells [101]. So, it is possible that *IRF2BP2* has a role in hair follicle pigmentation via regulation of other transcription factors. There is not much known about the relevance of *IRF2BP2* with human pigmentation. So, the functional relevance and expression patterns of this gene in differing hair color follicle or pigmentation in general needs to be understood.

A single factor analysis of variance was performed to identify genes that were significantly differing between blond Vs dark brown and red Vs dark brown. 5 genes seemed to show significance at certain comparisons. *KIT* (p= 0.0387), *PLXNC1* (p=0.00099*, Significant even after Bonferroni correction) and *PMEL* (p=0.0397) were significant between red and dark brown. *SLC16A1* (p=0.01402) and *TYRP1* (p=0.0085) were found to be significant between blond and dark brown.

Of these six genes, *KIT*, *PMEL*, *PLXNC1* and *TYRP1* have already been associated with pigmentation pathway. When looking at hair color in particular, *PMEL* (premelanosome protein) also known as *SILV* (silver locus homolog) has previously been associated with silver colored coat hair in ponies[102] and is regulated by *MITF*, a critical regulator for melanocyte development[103]. A SNP from *TYRP1* is a part of our HIrisPlex prediction system and was found associated with hair color. Also, from our previous chapters, it was noted that rs683 (an exonic SNP in *TYRP1*) improved the accuracy of hair color prediction when its interaction with rs12896399 (*SLC24A4*) was included. *TYRP1* act downstream in

the *MC1R* triggered pathway and are crucial in melanin synthesis. Differences at this gene has been associated with light- dark pigment variation in other species[86]. Mutations in *KIT* gene have previously been associated with a lack of melanocytes and it was shown that these mutations caused white coat color in pigs [104]. *KIT* and its ligand *KITLG* trigger the RAS/MAPK pathway which play a role in melanogenesis, melanosome transfer, etc.[105]. A study on profiling the aging process identified that *PLXNC1* (Plexin C1) was upregulated in graying hair than in pigmented hair [49]. Our results show that *PLXNC1* was upregulated in red hairs when compared to dark brown. *PLXNC1* codes for transmembrane receptors that regulate the adhesion of melanocytes and it are involved in the axon guidance pathway. *SLC16A1* was downregulated in both blond and red hairs. However, the fold change was only significant with blond hairs. Interestingly, *SLC16A1* (solute carrier family 16 member 1) was previously associated with being downregulated in lighter skinned chickens and there is no association of this gene thus far with human pigmentation [106].

The following genes- *KITLG, C10Orf11, MLANA, SLC16A9, SLC2A8, SLC30A1, TRPM* and *TYR,* showed increased expression in blond hairs and red hairs in comparison to dark brown. Studies on mice models show that the expression of *KITLG* (ligand for *KIT*, a tyrosine kinase receptor) is down-regulated in lighter pigmented hairs [51]. We observed an increase in expression in both blond and red hair when compared to dark brown hairs. *C10orf11* (chromosome 10 open reading frame 11) is a melanocyte differentiation gene, which causes albinism (autosomal-recessive, *OCA7*) in humans [107]. The results from our study show that C10orf11 is down-regulated in dark brown hairs and up- regulated in light colors. *MLANA* (chromosome 9) is a protein-coding gene which plays a role in the melanogenesis pathway. *PMEL* (chromosome 12) is a melanocyte protein that ensures formation of stage II melanosomes. Both these genes are localized in melanosomes and are both regulated by Microphthalmia-associated Transcription Factor (*MITF*) [103]. It is interesting to note that we observed both *MLANA* and *PMEL* to be up- regulated in blond

and red hairs. Skin color transcriptome profiles for chickens were analyzed for white vs black skin, and *SLC16A9* was highly expressed in darker colored skin[106].

Transient receptor potential melastatin (*TRPM*) is an ion channel that potentially controls calcium homeostasis in human melanocytes. Knockdown of *TRPM1* was shown to decrease tyrosinase activity and therefore melanin pigment. Downregulation of *TRPM1* has also been associated with spotting pattern in horses[108]. Our data, shows that *TRPM* was upregulated in both blond & red hairs and was down-regulated in dark brown hairs. *IRF2* belongs to the same family as *IRF4* which is associated with hair color[109]. Lower expression of *IRF4* has been associated with lighter hair color [34, 35, 53]. In our study, expression of *IRF2* was higher in blond hair when compared to dark brown hair. It is down-regulated in darker pigmented hair and up-regulated in lighter hair color. And since, *IRF2* and *IRF4* belong to the same family of interferon regulatory factors, they probably take up similar expression patterns for each hair color. Even though there was an observable fold change in *IRF2*, statistical calculations did not turn out significant.

PPF1BP1 (protein tyrosine phosphatase receptor type F polypeptide interacting protein binding protein 1) was downregulated in both blond and red hairs. *PPF1BP1* is involved in axon guidance, is upregulated by preimplantation factor (*PIF*)[110], and interacts with calcium binding proteins. An interesting observation is that, three of the genes- PPF1BP1 (downregulated in blond and red hairs), *PLXNC1* (upregulated in blond and red hairs), and PLXNB1 (no difference in expression) are associated with axon guidance pathway. *MIR29B2* (microRNA 29b-2) and *C10orf44* (Chromosome 10 open reading frame 44) showed higher expression in blond hair and lower expression in red hairs. Unlike *C10orf11*, there is no evidence of *C10orf44* playing a role in melanocyte differentiation yet. Roles of *PPF1BP1*, *MIR29B2* and *C10orf44* have also not been studied with relevance to pigmentation.

Although many of the candidate genes had fold difference between hair colors, many of them were not statistically significant. This is probably due to low power. In fact, some of these genes are major contributors to the pigmentation pathway especially, *KITLG* and *TYR*. Unlike model organisms, it is difficult to obtain samples from individuals with identical and controlled environmental conditions. This variability and the inherent differences between individuals could have attributed to insignificant results. This could potentially be overcome by increasing the sample size and also by confirming the differential expression using RNA-sequencing data. It is still interesting to note that some of the genes known to be involved with pigmentation pathway; like *KIT*, *PMEL*, *TYRP1*, *IRF2B2* and *PLXNC1;* were statistically significant. For the remainder of the genes, although some of them showed reasonable fold change difference, the results were not statistically significant (*p*-values range: 0.053-0.982). Again, the reason could be attributed to the innate difference that might arise between individuals, which is expected when dealing with higher order organisms, leading to high variability between the samples, therefore an increase in sample size could correct for this. All the genes were identified via RNA-sequencing bioinformatics analysis in our concurrent study on non-human primate pelage. Nevertheless, this pilot study was conducted on candidate genes and we plan to expand our study subjects and probe into more candidate genes for gene expression.

CHAPTER 6. CONCLUSIONS

The first goal of this study was to identify any potential epistatic interactions by implementing MDR and logistic regression methods between known pigmentation markers, and then see if they had additive effect on improving the prediction accuracy of categorical classification of hair color in the HIrisPlex system. For this study, we analyzed 872 samples at 61 markers and identified 3 interactions between genes, *SLC24A4* rs12896399 * *TYRP1* rs683 (*p*-val of interaction=0.019), *IRF4* rs12203592 * *TPCN2* rs35264875 (*p*-val of interaction= 1.499E-5), *ASIP* rs1015362 * *BNC2* rs10756819 (*p*-val of interaction= 0.017). Apart from the SNP-SNP interactions, we found that rs26722 (*SLC45A2*) also had a significant independent effect on hair color in this dataset. Hair color is more of a complex continuous trait and studies are ongoing to identify more genes and SNPs that could give a better understanding of pigmentation genetics at a continuous level. However, based on the results of this study, it is clear that SNP- SNP interactions should also be included in addition to variants with independent effects while building new prediction models.

The second goal of the study was to look at a panel of candidate genes that showed promise in an ongoing study on non-human primates. These candidate genes were assessed for differential gene expression in differing hair color follicles (within same individuals) using qPCR on our study set to examine their expression. Upon statistical analysis, 6 genes- *IRF2BP2 (p*= 0.0377*)*, *KIT (p*= 0.0387*)*, *PMEL (p*=0.0397*)*, *TYRP1 (p*=0.0085*)*, *SLC16A1 (p*=0.014*)* and *PLXNC1 (p*=0.00099**, **Significant after Bonferroni correction*)* were found to have significant differential gene expression between particular hair color categories.

Parallel studies on humans and non-human primates would also help us gain better understanding from an evolutionary perspective. When working with higher order organisms like humans, it is inevitable that we come across high variability between individuals. This can be reduced by increasing the sample size. Although some of the candidate genes used for this study have already been acknowledged for having a part in the melanogenesis pathway, there may be many more that need to be explored at the transcriptome level.

REFERENCES

REFERENCES

1. Butler, J.M., *The future of forensic DNA analysis.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1674).
2. MacDonald, F., *Kuwait has become the first country to make DNA testing mandatory for all residents*. 2015, Science alert.
3. Walsh, S., et al., *Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence.* Forensic Sci Int Genet, 2011. **5**(5): p. 464-71.
4. Kayser, M., *Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes.* Forensic Sci Int Genet, 2015. **18**: p. 33-48.
5. Walsh, S., et al., *IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information.* Forensic Sci Int Genet, 2011. **5**(3): p. 170-80.
6. Walsh, S., et al., *The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA.* Forensic Sci Int Genet, 2013. **7**(1): p. 98-115.
7. Ruiz, Y., et al., *Further development of forensic eye color predictive tests.* Forensic Sci Int Genet, 2013. **7**(1): p. 28-40.
8. Sochtig, J., et al., *Exploration of SNP variants affecting hair colour prediction in Europeans.* Int J Legal Med, 2015. **129**(5): p. 963-75.
9. Mushailov, V., et al., *Assay Development and Validation of an 8-SNP Multiplex Test to Predict Eye and Skin Coloration.* J Forensic Sci, 2015. **60**(4): p. 990-1000.
10. Park, H.Y., et al., *Cellular mechanisms regulating human melanogenesis.* Cell Mol Life Sci, 2009. **66**(9): p. 1493-506.

11. Sturm, R.A., N.F. Box, and M. Ramsay, *Human pigmentation genetics: the difference is only skin deep.* Bioessays, 1998. **20**(9): p. 712-21.
12. Schlake, T., *Determination of hair structure and shape.* Semin Cell Dev Biol, 2007. **18**(2): p. 267-73.
13. Slominski, A., et al., *Hair follicle pigmentation.* J Invest Dermatol, 2005. **124**(1): p. 13-21.
14. Clinics, A. *Know it all about the mood-swings of your hair!!The complete Hair Cycle*. 2013 11/22/2016]; Available from: https://akclinics.org/blog/hair-loss-articles/hair-growth-cycle/.
15. Cichorek, M., et al., *Skin melanocytes: biology and development.* Postepy Dermatol Alergol, 2013. **30**(1): p. 30-41.
16. Matama, T., A.C. Gomes, and A. Cavaco-Paulo, *Hair Coloration by Gene Regulation: Fact or Fiction?* Trends Biotechnol, 2015. **33**(12): p. 707-11.
17. 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, 2007. **80**(2): p. 241-52.
18. Branicki, W., U. Brudnik, and A. Wojas-Pelc, *Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype.* Ann Hum Genet, 2009. **73**(2): p. 160-70.
19. Visser, M., M. Kayser, and R.J. Palstra, *HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter.* Genome Res, 2012. **22**(3): p. 446-55.
20. Sturm, R.A. and T.N. Frudakis, *Eye colour: portals into pigmentation genes and ancestry.* Trends Genet, 2004. **20**(8): p. 327-32.
21. Frudakis, T., et al., *Sequences associated with human iris pigmentation.* Genetics, 2003. **165**(4): p. 2071-83.
22. Sulem, P., et al., *Genetic determinants of hair, eye and skin pigmentation in Europeans.* Nat Genet, 2007. **39**(12): p. 1443-52.

23. 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, 2008. **123**(2): p. 177-87.
24. Valenzuela, R.K., et al., *Predicting phenotype from genotype: normal pigmentation.* J Forensic Sci, 2010. **55**(2): p. 315-22.
25. Eriksson, N., et al., *Web-based, participant-driven studies yield novel genetic associations for common traits.* PLoS Genet, 2010. **6**(6): p. e1000993.
26. Kayser, M.H., F. Liu, and A. Hofman, *Method for prediction of human iris color*. 2011, Google Patents.
27. Pospiech, E., et al., *The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction.* Forensic Sci Int Genet, 2014. **11**: p. 64-72.
28. Freire-Aradas, A., et al., *Exploring iris colour prediction and ancestry inference in admixed populations of South America.* Forensic Sci Int Genet, 2014. **13**: p. 3-9.
29. Sturm, R.A., et al., *Genetic association and cellular function of MC1R variant alleles in human pigmentation.* Ann N Y Acad Sci, 2003. **994**: p. 348-58.
30. Han, J., et al., *A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation.* PLoS Genet, 2008. **4**(5): p. e1000074.
31. Liu, F., et al., *Eye color and the prediction of complex phenotypes from genotypes.* Curr Biol, 2009. **19**(5): p. R192-3.
32. Sturm, R.A., *Molecular genetics of human pigmentation diversity.* Hum Mol Genet, 2009. **18**(R1): p. R9-17.
33. Mengel-From, J., et al., *Genetic determinants of hair and eye colours in the Scottish and Danish populations.* BMC Genet, 2009. **10**: p. 88.
34. Praetorius, C., et al., *A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway.* Cell, 2013. **155**(5): p. 1022-33.

35. Adhikari, K., et al., *A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features.* Nat Commun, 2016. **7**: p. 10815.

36. Brilliant, M.H. *Gene Polymorphism and Human Pigmentation* 2008 11/22/2016]; Available from: https://www.ncjrs.gov/pdffiles1/nij/grants/223980.pdf.

37. 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, 2005. **116**(5): p. 402-6.

38. Lao, O., et al., *Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms.* Ann Hum Genet, 2007. **71**(Pt 3): p. 354-69.

39. Gudbjartsson, D.F., et al., *ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma.* Nat Genet, 2008. **40**(7): p. 886-91.

40. Edwards, M., et al., *Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation.* PLoS Genet, 2010. **6**(3): p. e1000867.

41. Hart, K.L., et al., *Improved eye- and skin-color prediction based on 8 SNPs.* Croat Med J, 2013. **54**(3): p. 248-56.

42. Liu, J.Z., et al., *A versatile gene-based test for genome-wide association studies.* Am J Hum Genet, 2010. **87**(1): p. 139-45.

43. 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, 2013. **132**(2): p. 147-58.

44. Maronas, O., et al., *Development of a forensic skin colour predictive test.* Forensic Sci Int Genet, 2014. **13**: p. 34-44.

45. Jacobs, L.C., et al., *A Genome-Wide Association Study Identifies the Skin Color Genes IRF4, MC1R, ASIP, and BNC2 Influencing Facial Pigmented Spots.* J Invest Dermatol, 2015. **135**(7): p. 1735-42.

46. Rutberg, S.E., et al., *Differences in expression of specific biomarkers distinguish human beard from scalp dermal papilla cells.* J Invest Dermatol, 2006. **126**(12): p. 2583-95.

47. Rosengren Pielberg, G., et al., *A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse.* Nat Genet, 2008. **40**(8): p. 1004-9.

48. Chiaverini, C., et al., *Cystinosin is a melanosomal protein that regulates melanin synthesis.* FASEB J, 2012. **26**(9): p. 3779-89.

49. Peters, E.M., et al., *Profiling mRNA of the graying human hair follicle constitutes a promising state-of-the-art tool to assess its aging: an exemplary report.* J Invest Dermatol, 2013. **133**(5): p. 1150-60.

50. Korberg, I.B., *A Simple Repeat Polymorphism in the MITF-M Promoter Is a Key Regulator of White Spotting in Dogs.* PLoS One, 2014. **9**(8).

51. Guenther, C.A., et al., *A molecular basis for classic blond hair color in Europeans.* Nat Genet, 2014. **46**(7): p. 748-52.

52. Wu, Z., et al., *Identification of differentially expressed miRNAs between white and black hair follicles by RNA-sequencing in the goat (Capra hircus).* Int J Mol Sci, 2014. **15**(6): p. 9531-45.

53. Visser, M., R.J. Palstra, and M. Kayser, *Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter.* Hum Mol Genet, 2015. **24**(9): p. 2649-61.

54. Ahmed, M.I., et al., *MicroRNA-214 controls skin and hair follicle development by modulating the activity of the Wnt pathway.* J Cell Biol, 2014. **207**(4): p. 549-67.

55. Tian, X., et al., *Molecular cloning, mRNA expression and tissue distribution analysis of Slc7a11 gene in alpaca (Lama paco) skins associated with different coat colors.* Gene, 2015. **555**(2): p. 88-94.

56. Walsh, S., et al., *Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage.* Forensic Sci Int Genet, 2014. **9**: p. 150-61.

57. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.

58. Untergasser, A., et al., *Primer3Plus, an enhanced web interface to Primer3.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W71-4.

59. Aranyi, T., et al., *The BiSearch web server.* BMC Bioinformatics, 2006. **7**: p. 431.

60. Vallone, P.M. and J.M. Butler, *AutoDimer: a screening tool for primer-dimer and hairpin structures.* Biotechniques, 2004. **37**(2): p. 226-31.

61. Keating, B., et al., *First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip.* Int J Legal Med, 2013. **127**(3): p. 559-72.

62. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

63. Shriver), A.F.w.c.b.M.D., *Molecular Photofitting: Predicting Ancestry and Phenotype Using DNA*. 2010: Elsevier.

64. Liu, F., et al., *Detecting low frequent loss-of-function alleles in genome wide association studies with red hair color as example.* PLoS One, 2011. **6**(11): p. e28145.

65. Kastelic, V. and K. Drobnic, *A single-nucleotide polymorphism (SNP) multiplex system: the association of five SNPs with human eye and hair color in the Slovenian population and comparison using a Bayesian network and logistic regression model.* Croat Med J, 2012. **53**(5): p. 401-8.

66. Sulem, P., et al., *Two newly identified genetic determinants of pigmentation in Europeans.* Nat Genet, 2008. **40**(7): p. 835-7.

67. Hudjashov, G., R. Villems, and T. Kivisild, *Global patterns of diversity and selection in human tyrosinase gene.* PLoS One, 2013. **8**(9): p. e74307.

68. Jacobs, L.C., et al., *IRF4, MC1R and TYR genes are risk factors for actinic keratosis independent of skin color.* Hum Mol Genet, 2015. **24**(11): p. 3296-303.

69. 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, 2008. **82**(2): p. 411-23.

70. 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, 2008. **128**(12): p. 2807-14.

71. Dimisianos, G., et al., *A study of a single variant allele (rs1426654) of the pigmentation-related gene SLC24A5 in Greek subjects.* Exp Dermatol, 2009. **18**(2): p. 175-7.

72. Siiskonen, S.J., et al., *A Genome-Wide Association Study of Cutaneous Squamous Cell Carcinoma among European Descendants.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(4): p. 714-20.

73. Flanagan, N., et al., *Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation.* Hum Mol Genet, 2000. **9**(17): p. 2531-7.

74. Jonnalagadda, M., et al., *Association of genetic variants with skin pigmentation phenotype among populations of west Maharashtra, India.* Am J Hum Biol, 2016. **28**(5): p. 610-8.

75. Branicki, W., et al., *Model-based prediction of human hair color using DNA variants.* Hum Genet, 2011. **129**(4): p. 443-54.

76. Liu, F., et al., *Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up.* Hum Genet, 2015. **134**(8): p. 823-35.

77. Hughes, T., et al., *Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus.* Arthritis Rheum, 2012. **64**(2): p. 485-92.

78. Konoenko, I. *Estimating attributes: analysis and extensions of Relief.* in *European Conference on Machine Learning*. 1994.

79. Robnik-Sikonja, M.K., I, *Theoretical and empirical analysis of ReliefF and RReliefF.* Machine Learning, 2003. **53**: p. 23–69.

80. Pospiech, E., et al., *Gene-gene interactions contribute to eye colour variation in humans.* J Hum Genet, 2011. **56**(6): p. 447-55.

81. Walsh, S., et al., *Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage.* Forensic Science International-Genetics, 2014. **9**: p. 150-161.

82. Team, R. *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA*. 2015; Available from: http://www.rstudio.com/.

83. Croissant, Y. *Estimation of multinomial logit models in R : The mlogit Packages*. 2013;Available from: https://cran.r-project.org/web/packages/mlogit/index.html

84. Tobias Sing, O.S., Niko Beerenwinkel, Thomas Lengauer. *Visualizing the Performance of Scoring Classifiers*. 2015; Available from: http://rocr.bioinf.mpi-sb.mpg.de/.

85. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.

86. Brenda J. Bradley, M.S.G., Anja Widdig &Nicholas I. Mundy, *Coat Color Variation and Pigmentation Gene Expression in Rhesus Macaques (Macaca mulatta).* J Mammal Evol, 2013. **20**: p. 8.

87. Schmittgen, T.D. and K.J. Livak, *Analyzing real-time PCR data by the comparative C(T) method.* Nat Protoc, 2008. **3**(6): p. 1101-8.

88. Moore, J.H., et al., *A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility.* J Theor Biol, 2006. **241**(2): p. 252-61.

89. Camargo, A., et al., *Permutation - based statistical tests for multiple hypotheses.* Source Code Biol Med, 2008. **3**: p. 15.

90. *Introduction to SAS. UCLA: Statistical Consulting Group.* . 11/22/2016]; Available from: http://www.ats.ucla.edu/stat/sas/notes2/.

91. Donnelly, M.P., et al., *A global view of the OCA2-HERC2 region and pigmentation.* Hum Genet, 2012. **131**(5): p. 683-96.

92. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* Radiology, 1982. **143**(1): p. 29-36.

93. Saternus, R., et al., *A closer look at evolution: Variants (SNPs) of genes involved in skin pigmentation, including EXOC2, TYR, TYRP1, and DCT, are associated with 25(OH)D serum concentration.* Endocrinology, 2015. **156**(1): p. 39-47.

94. Steiner, C.C., J.N. Weber, and H.E. Hoekstra, *Adaptive variation in beach mice produced by two interacting pigmentation genes.* PLoS Biol, 2007. **5**(9): p. e219.

95. Kocarnik, J.M., et al., *Pleiotropic and sex-specific effects of cancer GWAS SNPs on melanoma risk in the population architecture using genomics and epidemiology (PAGE) study.* PLoS One, 2015. **10**(3): p. e0120491.

96. Lu, D., et al., *Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor.* Nature, 1994. **371**(6500): p. 799-802.

97. Galvan, I. and C. Alonso-Alvarez, *The expression of melanin-based plumage is separately modulated by exogenous oxidative stress and a melanocortin.* Proc Biol Sci, 2009. **276**(1670): p. 3089-97.

98. Visser, M., R.J. Palstra, and M. Kayser, *Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene.* Hum Mol Genet, 2014. **23**(21): p. 5750-62.

99. Praetorius, C., R.A. Sturm, and E. Steingrimsson, *Sun-induced freckling: ephelides and solar lentigines.* Pigment Cell Melanoma Res, 2014. **27**(3): p. 339-50.

100. Carneiro, F.R., et al., *Interferon regulatory factor 2 binding protein 2 is a new NFAT1 partner and represses its transcriptional activity.* Mol Cell Biol, 2011. **31**(14): p. 2889-901.

101. Lee, Y.S., et al., *Downregulation of NFAT2 promotes melanogenesis in B16 melanoma cells.* Anat Cell Biol, 2010. **43**(4): p. 303-9.

102. Reissmann, M., J. Bierwolf, and G.A. Brockmann, *Two SNPs in the SILV gene are associated with silver coat colour in ponies.* Anim Genet, 2007. **38**(1): p. 1-6.

103. Hoek, K.S., et al., *Novel MITF targets identified using a two-step DNA microarray strategy.* Pigment Cell Melanoma Res, 2008. **21**(6): p. 665-76.

104. Johansson Moller, M., et al., *Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor.* Mamm Genome, 1996. **7**(11): p. 822-30.

105. Picardo, M. and G. Cardinali, *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, 2011. **131**(6): p. 1182-5.

106. Zhang, J., et al., *Skin transcriptome profiles associated with skin color in chickens.* PLoS One, 2015. **10**(6): p. e0127301.

107. Gronskov, K., et al., *Mutations in c10orf11, a melanocyte-differentiation gene, cause autosomal-recessive albinism.* Am J Hum Genet, 2013. **92**(3): p. 415-21.

108. Devi, S., et al., *Calcium homeostasis in human melanocytes: role of transient receptor potential melastatin 1 (TRPM1) and its regulation by ultraviolet light.* Am J Physiol Cell Physiol, 2009. **297**(3): p. C679-87.

109. Yanai, H., H. Negishi, and T. Taniguchi, *The IRF family of transcription factors: Inception, impact and implications in oncogenesis.* Oncoimmunology, 2012. **1**(8): p. 1376-1386.

110. Duzyj, C.M., et al., *PreImplantation factor (PIF*) promotes embryotrophic and neuroprotective decidual genes: effect negated by epidermal growth factor.* J Neurodev Disord, 2014. **6**(1): p. 36.

APPENDICES

Appendix A SNaPshot Multiplex Assay Primer Information

Assay	SNP	Gene	Forward and Reverse primers	SBE primer
Philips-Skin	rs10777129	KITLG	TCAAGTGACTCAAAGTCATACAGTTT	tttTAGCCTCAAATTAGTTTTTGTATAGTA
			TGAGCCTCAAATTAGTTTTTGT	
	rs6058017	ASIP	GTGCTCAGCCTCAACTGCT	ttttttttttttttttttttttttTCCGAAGCCCTGCC
			AGTCCCATCTCCTGGAACCT	
	rs13289	SLC45A2	CAGGTCACACCCTTCTTCAAA	ttttttttttttttttttttTTCGAGGAGAAATATCAGGGC
			AAACCCAGAATTATCTCATTGTAAA	
	rs1408799	TYRP1	TCAAAATCAAAACTGGTTCATCC	ttttttttttttttttttttttttttttttTCGGAGCACATGGTCA
			GTGCTATGAGGACAGGACCAT	
	rs3829241	TPCN2	CTGCTCCCTTGTTCTTGTGG	ttttttttttttttttttttttttttttttttttTCTGGAGGAGCCCG
			GCTGGCTCAGCCTCTCTGT	
	rs1448484	OCA2	GCAACCACACAGAACACAGC	ttttttttTTGATCTGAGGAAGAAACATGAGTA
			TCTCCTTCCAAGCCTTCTGA	
ComPlex	rs11636232	HERC2	TGCTAAGAACCACACAGCACCT	TTTTTGTTCCCCTCCGATTAA
			CATTGAAGGCGCAAAAGTC	
	rs26722	SLC45A2	TGCCAGCTCTGGATTTACG	TTTTTTTTTTTTTTTTTTTTTTTTTTTACGTAACCATTTTTAACTTTCT
			ACAGCAAACCCCTCAGGAC	
	rs7183877	HERC2	TGGCTCTCTGTGTCTGATCC	TTTTTTTTTAAGCAGTATACATTTAGAAATGGT
			TGTCTCATGGGTAGTAATCAAAGAA	
	rs12592730	HERC2	aagacagaaaagctgccaaga	TTTTTTTTTTTTTTTTTTTTTTCTGTTATTGGCTGGAGTGTT
			tgctgttattggctggagtg	

	rs916977	HERC2	TTGTTTTGGCAAACTCCACA	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGATGCAGTTTGAGTAGA
			CCTTGGTGATGGCTTAGGAG	
	rs1015362	ASIP	CAAATAGTCCCGACCAGGAG	TTTTTTTTTTTTTTTTTTTCAGGAGATGAAAACATCTCA
			TAACCCCTCCTTTTCTTTGC	
	rs12931267	FANCA	GAAGTTCCCAGTTCTCCTCCT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGTGATGGCTGTGCTTCT
			atttcctaggttggggttgg	
	rs35264875	TPCN2	TTCTCAACTGCGTCTTCATTGTG	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAAAGACCTTGAGCAGCA
			AGATGGGCCAGCACAGAG	
	rs4778241	OCA2	gttcactcctgcatcaccaa	TTTTTAATTGTTGGCTGGTAGTTGCAATT
			ccatttgcgtgtagggtttt	
	rs8024968	OCA2	GGCCCTATGCACACATTTTC	TTTTTTTTTTTTTTGGAGAGTACAGATTCACAGACTT
			TTCACCTTGGTGCCTTAGAT	
	rs1375164	OCA2	GGTCATATCCCAGGGCAGAG	TTTTTTTTTTTTTTTTTTTTTTTTTTTCCCTGTCCTGTTGTTGTCA
			ACCCCAAATTTCCATCACAA	
	rs4778138	OCA2	gagaggaggaaaatctgcacac	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCACTGATTTAGCTGTGTTCTG
			gctctccttttgataccagca	
	rs7495174	OCA2	AGGGGGCGACTTTTTCCT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGCAAGTTCCCCTAAAGGT
			AGGCAAGTTCCCCTAAAGGT	
	rs4778232	OCA2	GAGGCAGATCCACCATTGTC	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGATCTAGGGATGAGGAA

			GAACCAAGGGATCTAGGGATG	
HIrisPlex	N29insA	MC1R	GCAGGGATCCCAGAGAAGAC	CCCCAGCTGGGGCTGGCTGCCAA
			TCAGAGATGGACACCTCCAG	
	rs11547464	MC1R	GTCCAGCCTCTGCTTCCTG	ttttttttttttGCCATCGCCGTGGACC
			AGCGTGCTGAAGACGACAC	
	rs885479	MC1R		ttttttttttttttttttGATGGCCGCAACGGCT
	rs1805008	MC1R		ttttttttttttACAGCATCGTGACCCTGCCG
	rs1805005	MC1R	CTGGTGAGCTTGGTGGAGA	ttttttttttttttTGGTGGAGAACGCGCTGGTG
			TCCAGCAGGAGGATGACG	
	rs1805006	MC1R		ttttttttttttttttttttCTGCCTGGCCTTGTCGGA
	rs1805007	MC1R		ttttttttttttttttttttttttttCTCCATCTTCTACGCACTG
	rs1805009	MC1R	CAAGAACTTCAACCTCTTTCTCG	ttttttttttttttttttttttttttttttttATCTGCAATGCCATCATC
			CACCTCCTTGAGCGTCCTG	
	Y152OCH	MC1R		ttttttttttttttttttttttttttttttttCATCTTCTACGCACTGCGCTA
	rs2228479	MC1R		ttttttttttttttttttttttttttttttttttttCTGGTGAGCGGGAGCAAC
	rs1110400	MC1R		ttttttttttttttttttttttttttttttttCTTCTACGCACTGCGCTACCACAGCA
	rs28777	SLC45A2	TACTCGTGTGGGAGTTCCAT	ttttttttttttttttttttttttttttttttttttttttCATGTGATCCTCACAGCAG
			TCTTTGATGTCCCCTTCGAT	
	rs16891982	SLC45A2	TCCAAGTTGTGCTAGACCAGA	ttttttttttttttttttttttttttttttttttttttttttttttttAAACACGGAGTTGATGCA
			CGAAAGAGGAGTCGAGGTTG	
	rs12821256	KITLG	ATGCCCAAAGGATAAGGAAT	ttttttttttttttttttttttttttttttttttttttttGGAGCCAAGGGCATGTTACTACGGCAC

			GGAGCCAAGGGCATGTTACT	
	rs4959270	EXOC2	Tgagaaatctacccccacga	ttttttttttttttttttttttttttttttttttttttttttttttttGGAACACATCCAAACTATGACACTATG
			GTGTTCTTACCCCCTGTGGA	
	rs12203592	IRF4	AGGGCAGCTGATCTCTTCAG	ttttttttttttttttttttttttttttttttttttttttttttttttTCCACTTTGGTGGGTAAAAGAAGG
			GCTTCGTCATATGGCTAAACCT	
	rs1042602	TYR	CAACACCCATGTTTAACGACA	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttTCAATGTCTCTCCAGATTTCA
			GCTTCATGGGCAAAATCAAT	
	rs1800407	OCA2	AAGGCTGCCTCTGTTCTACG	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttGCATACCGGCTCTCCC
			CGATGAGACAGAGCATGATGA	
	rs2402130	SLC24A4	acctgtctcacagtgctgct	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttTGAACCATACGGAGCCCGTG
			TTCACCTCgatgacgatgat	
	rs12913832	HERC2	TCAACATCAGGGTAAAAATCATGT	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttTAGCGTGCAGAACTTGACA
			GGCCCCTGATGATGATAGC	
	rs2378249	PIGU	CGCATAACCCATCCCTCTAA	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttCCACACCTCTCCTCAGCCCA
			CATTGCTTTTCAGCCCACAC	
	rs12896399	SLC24A4	CTGGCGATCCAATTCTTTGT	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttTCTTTAGGTCAGTATATTTTGGG
			GACCCTGTGTGAGACCCAGT	
	rs1393350	TYR	TTTCTTTATCCCCCTGATGC	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttCATTTGTAAAAGACCACACAGATTT
			GGGAAGGTGAATGATAACACG	

	rs683	TYRP1	CACAAAACCACCTGGTTGAA	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttGCTTTGAAAAGTATGCCTAGAACTTTAAT
			TGAAAGGGTCTTCCCAGCTT	
HIrisPlex-S	rs2238289	HERC2	GGAACATGAAGATTTCCCAGT	TTTTTTTTTTTTTTTTTTTTGAGATTGGAAGATTGGAGCC
			CTGATTCAGGTCTGCTGTCACT	
	rs6497292	HERC2	TCTGCTGTAGAACCAATGTCC	TTTTTTTTTTTTTTTTTTTTTTTTTTGTCTCCTGTGTCTTCATCCT
			TGTAGCTCCATTGGTGTCCTC	
	rs1129038	HERC2	ATGTCGACTCCTTTGCTTCG	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGCCAGGCAGCAGAGC
			ACACCAGGCAGCCTACAGTC	
	rs1667394	HERC2	CAGCTGTAGAGAGAGACTTTGAGG	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAGCAATTCAAAACGTGCATA
			CACCATTAAGACGCAGCAAT	
	rs3114908	ANKRD11	CAGAACACAGCCACACCCTA	TTTTTTTTTTTCGTCCCTTTCTGCCACTAC
			CATAAAGGGGTCACCAGCAA	
	rs10756819	BNC2	AAGCTCATGTTTCCAAATATGCTA	TTTTTTTTTTTTTTGGACCAGTTATTTTGGGTTTGGA
			CATCCCGTCATGACTAGAAAAA	
	rs17128291	SLC24A4	CCAGCACTGCCAAAATAACA	TTTTTTTTTTTTTTTTTTTTTCAATGTGCACTGGATTAAAAGTC
			CTCTTTGGACCCATCACCTC	
	rs1800414	OCA2	GCTGCAGGAGTCAGAAGGTT	TTTTTTTTTTTCAGAATCCCGTCAGATATCCTA
			GGGACAAACGAATTGAGGAA	
	rs1126809	TYR	TGTTTCTTAGTCTGAATAACCTTTTCC	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTATTTTTGAGCAGTGGCTCC
			GGTGCATTGGCTTCTGGATA	

	rs1470608	OCA2	TTTCTTGTGTTAACTGTCCTTACAAA	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTCGAACAACTAGTTTCATTACAGA
			GGAAAATATGTTAGGGTTGATGG	
	rs1426654	SLC24A5	TTCAGCCCTTGGATTGTCTC	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTCTCAGGATGTTGCAGGC
			TGAGTAAGCAAGAAGTATAAGGAGCA	
	rs6119471	ASIP	GCAGGAGAATTGCTGGAACT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAGGAAGAGTGAAAATGCGTAA
			AACCCGAAGGAAGAGTGAAAA	
	rs1545397	OCA2	GGTATAGGATTATTTGGGGAATGA	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTACAACTTTGTGAATATACTAAAATAC
			AAATGGAGATATAGAATTCACACAACA	
	rs6059655	RALY	GTGAGGAAATCGAGGCTCAG	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGAAAGGCTGAGTGGC
			AGGAGAAAGCTGCAGATCCA	
	rs12441727	OCA2	AGTGGGAAGAGACAGCTCCA	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGCTCAGTGTGGCCTT
			ACAATCCTGGGAGGTACACG	
	rs3212355	MC1R	GAGTGAACCCAGGAAGATGC	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCGAAGCCCAGCAGG
			CATCAAAGGCAGACCTCTCG	
	rs8051733	DEF8	AGGCGGTGGTCTCTCTCTC	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCACCCTGCCTGTCTCG
			TTGCAACAGGAGGGTCTAGG	

Appendix B PLINK Interaction Results

Listing all the interactions that were identified using BLR (PLINK) and are statistically significant (*p*-val<0.05). The ones highlighted in yellow are the interactions that were significant even after Bonferroni correction. Traits are classified as (Cases/ Controls) and OR represents odds ratio of 'cases'. STAT represents Chi- square (X^2) test statistic at 1 degree of freedom (df)

Trait classification	CHR1	SNP1	CHR2	SNP2	OR_INT	STAT	P-Val
Black/ Non- black	6	rs12203592	6	rs4959270	6.924	20.31	6.58E-06
	6	rs12203592	11	rs35264875	3.703	18.74	1.5E-05
	6	rs12203592	15	rs11636232	5.289	16.03	6.22E-05
	5	rs13289	16	rs885479	0.1637	15.89	6.72E-05
	6	rs12203592	15	rs1667394	0.258	15.67	7.53E-05
	6	rs12203592	11	rs3829241	3.418	14.75	0.000123
	6	rs12203592	15	rs2238289	0.2541	13.44	0.000246
	6	rs12203592	15	rs1129038	0.2937	13.06	0.000302
	11	rs3829241	11	rs35264875	2.582	12.86	0.000336
	12	rs10777129	15	rs1426654	0.3589	12.63	0.000379
	6	rs4959270	11	rs35264875	2.591	12.38	0.000434
	6	rs12203592	16	rs1805005	13.62	12.15	0.000492
	15	rs1667394	15	rs1426654	0.353	11.89	0.000565
	6	rs12203592	15	rs1375164	0.2895	11.56	0.000673
	6	rs12203592	9	rs683	0.3507	11.51	0.000693
	5	Rs16891982	6	rs4959270	0.3888	11.45	0.000714
	6	rs12203592	15	rs4778241	0.3014	11.19	0.00082
	5	Rs16891982	12	rs10777129	0.4083	10.89	0.000966
	15	rs4778138	15	rs1426654	0.3915	10.65	0.001102
	15	rs6497292	15	rs1426654	0.4105	10.27	0.00135
	6	rs12203592	11	rs1042602	2.555	9.737	0.001806

	15	rs1470608	15	rs1426654	0.3877	9.621	0.001924
	6	rs12203592	11	rs1126809	4.245	9.505	0.002049
	6	rs12203592	9	rs1408799	0.409	9.185	0.002441
	6	rs4959270	15	rs1129038	0.3703	9.013	0.002681
	6	rs12203592	16	rs2228479	3.125	8.999	0.002701
	11	rs1042602	11	rs1126809	5.272	8.947	0.002779
	15	rs1375164	15	rs1426654	0.4014	8.826	0.00297
	15	rs1667394	16	rs885479	0.3593	8.748	0.003099
	6	rs4959270	11	rs3829241	2.515	8.657	0.003258
	15	rs8024968	15	rs1426654	0.4299	8.19	0.004212
	6	rs12203592	15	rs4778138	0.2822	7.879	0.005002
	11	rs1126809	16	rs8051733	0.2973	7.868	0.005032
	11	rs3829241	16	rs1805005	4.58	7.789	0.005257
	5	rs13289	6	rs4959270	0.4813	7.268	0.00702
	6	rs4959270	16	rs1805005	7.933	7.241	0.007125
	6	rs12203592	15	rs1470608	0.3401	7.052	0.007918
	11	rs1042602	15	rs1129038	0.3972	6.941	0.008423
	11	rs35264875	15	rs1470608	0.4159	6.925	0.008502
	16	rs1805007	20	rs6058017	15.3	6.871	0.008761
	5	Rs16891982	6	rs12203592	0.2926	6.865	0.008788
	15	rs2238289	15	rs1426654	0.5021	6.802	0.009104
	9	rs1408799	16	rs1805005	0.2051	6.715	0.009561
	12	rs10777129	16	rs2228479	3.832	6.618	0.0101
	6	rs12203592	11	Rs1393350	2.539	6.587	0.01027
	11	rs35264875	15	rs1375164	0.4796	6.543	0.01053
	11	Rs1393350	16	rs8051733	0.3723	6.105	0.01348
	6	rs12203592	15	rs8024968	0.2365	6.094	0.01356
	15	rs7495174	20	rs6119471	0.1777	6.07	0.01375

	9	rs683	11	rs1126809	0.3268	6.024	0.01411
	9	rs683	11	rs3829241	0.4548	6.01	0.01422
	5	rs13289	9	rs10756819	1.889	5.987	0.01441
	5	rs13289	15	rs1426654	0.5287	5.953	0.01469
	11	rs35264875	15	rs11636232	2.477	5.884	0.01528
	11	rs1042602	11	Rs1393350	3.904	5.872	0.01538
	15	rs8024968	15	rs6497292	0.4786	5.832	0.01573
	11	rs35264875	16	rs1805005	3.827	5.75	0.01649
	15	rs7495174	15	rs1426654	0.447	5.732	0.01666
	11	rs3829241	14	rs17128291	3.102	5.63	0.01766
	15	rs12592730	15	rs1426654	0.4416	5.623	0.01773
	16	rs2228479	16	rs1805008	21.88	5.621	0.01774
	9	rs683	16	rs1805005	0.2657	5.598	0.01798
	15	rs1800414	16	rs2228479	10.76	5.449	0.01957
	6	rs12203592	9	rs10756819	0.4593	5.428	0.01982
	11	rs35264875	16	rs2228479	2.283	5.4	0.02014
	6	rs4959270	16	rs8051733	0.5469	5.383	0.02034
	6	rs4959270	20	rs2378249	0.3958	5.339	0.02085
	6	rs4959270	15	rs1375164	0.5442	5.321	0.02107
	9	rs10756819	11	rs1126809	0.3686	5.316	0.02113
	14	rs2402130	15	rs1545397	0.3869	5.245	0.022
	15	rs1448484	15	rs7495174	0.4674	5.141	0.02337
	15	rs12441727	20	rs1015362	0.3857	5.139	0.0234
	6	rs4959270	15	rs1667394	0.5259	5.119	0.02366
	12	rs10777129	15	rs1448484	0.5401	5.116	0.0237
	9	rs683	16	rs8051733	1.925	5.115	0.02371
	15	rs4778138	16	rs885479	0.4422	5.084	0.02414
	15	rs8024968	16	rs8051733	1.916	5.081	0.02419

	5	Rs16891982	15	rs1426654	0.4854	5	0.02535
	15	rs1375164	15	rs1129038	2.403	4.828	0.02801
	6	rs12203592	15	rs12441727	0.2069	4.809	0.02831
	16	rs8051733	20	rs6119471	4.214	4.719	0.02984
	9	rs683	15	rs1375164	1.974	4.679	0.03054
	14	Rs12896399	15	rs4778138	0.5125	4.663	0.03082
	9	rs1408799	15	rs11636232	0.3698	4.66	0.03088
	6	rs12203592	15	rs6497292	0.2242	4.566	0.0326
	11	rs3829241	20	rs2378249	0.3106	4.564	0.03266
	14	rs2402130	15	rs4778241	0.5473	4.533	0.03325
	15	rs12592730	20	rs1015362	0.3053	4.515	0.03361
	9	rs683	16	rs1805008	0.1579	4.484	0.03422
	11	rs1042602	15	rs1545397	0.2181	4.47	0.0345
	5	Rs16891982	16	rs8051733	1.711	4.375	0.03648
	14	rs2402130	14	rs17128291	2.577	4.373	0.03651
	9	rs1408799	20	rs6119471	0.327	4.348	0.03706
	11	rs35264875	15	rs8024968	0.4363	4.341	0.0372
	15	rs1448484	15	rs1426654	0.5015	4.321	0.03764
	16	rs1805005	16	rs1805008	16.2	4.318	0.03771
	6	rs4959270	16	rs2228479	2.436	4.317	0.03773
	5	rs13289	15	rs1800414	0.2259	4.296	0.03819
	14	rs17128291	16	rs885479	4.25	4.265	0.03891
	11	rs1042602	20	rs2378249	0.2804	4.262	0.03898
	12	rs10777129	15	rs1667394	0.5513	4.233	0.03965
	11	Rs1393350	16	rs1805005	3.903	4.216	0.04005
	16	rs8051733	20	rs2378249	2.011	4.197	0.0405
	9	rs10756819	11	rs3829241	0.5176	4.187	0.04075
	14	rs2402130	15	rs7495174	0.4505	4.185	0.04078

	6	rs4959270	15	rs1448484	0.4532	4.184	0.04081
	20	rs6059655	20	rs6119471	26.8	4.163	0.04132
	6	rs12203592	16	rs1805008	5.944	4.132	0.04207
	12	rs10777129	15	rs4778138	0.5756	4.121	0.04235
	11	rs3829241	15	rs4778241	0.4927	4.111	0.0426
	9	rs10756819	20	rs2378249	1.943	4.11	0.04264
	15	rs1470608	15	rs1375164	1.883	4.107	0.04271
	5	rs13289	15	rs12592730	0.5056	4.103	0.04282
	12	rs10777129	15	rs1800414	7.257	4.097	0.04296
	9	rs1408799	15	rs1545397	2.743	4.09	0.04315
	15	rs1448484	15	rs7183877	0.4463	4.084	0.0433
	6	rs4959270	9	rs1408799	0.5845	4.076	0.04351
	11	rs1126809	16	rs1805005	6.147	4.072	0.0436
	16	rs3114908	16	rs885479	2.616	4.005	0.04537
	6	rs4959270	11	rs1042602	1.894	3.977	0.04612
	12	rs10777129	20	rs6059655	4.947	3.948	0.04691
	5	rs13289	15	rs1545397	0.5268	3.894	0.04845
	15	rs7495174	20	rs1015362	0.4298	3.874	0.04903
Black/ Dark brown	6	rs12203592	6	rs4959270	0.1572	17.88	2.36E-05
	6	rs12203592	11	rs35264875	3.556	15.67	7.56E-05
	5	rs13289	16	rs885479	0.2129	11.41	0.00073
	6	rs12203592	16	rs1805005	12.04	11.21	0.000814
	6	rs12203592	11	rs3829241	3.001	10.83	0.000999
	6	rs12203592	15	rs1667394	0.3331	9.475	0.002083
	6	rs4959270	11	rs35264875	0.4294	9.357	0.002222
	6	rs12203592	15	rs11636232	3.685	9.104	0.00255
	12	rs10777129	15	rs1426654	0.4095	8.98	0.002729
	6	rs12203592	15	rs2238289	0.3198	8.714	0.003157

	5	Rs16891982	6	rs4959270	2.338	8.699	0.003183
	5	Rs16891982	12	rs10777129	0.4364	8.676	0.003224
	6	rs12203592	15	rs12913832	0.3613	8.522	0.003509
	11	rs1042602	11	rs1126809	5.361	8.429	0.003693
	12	rs10777129	15	rs1448484	0.3623	8.372	0.00381
	11	rs3829241	16	rs1805005	5.455	7.836	0.00512
	11	rs1042602	15	rs12913832	0.4174	7.092	0.007743
	11	rs1042602	11	Rs1393350	4.64	7.039	0.007974
	6	rs12203592	15	rs4778241	0.3684	6.947	0.008398
	6	rs12203592	11	rs1042602	2.3	6.88	0.008716
	6	rs4959270	11	rs3829241	0.4355	6.791	0.009164
	9	rs683	11	rs3829241	0.4084	6.554	0.01046
	11	rs1126809	16	rs8051733	0.3103	6.509	0.01073
	15	rs12441727	20	rs1015362	0.3287	6.42	0.01129
	9	rs1408799	16	rs1805005	0.1992	6.399	0.01142
	6	rs4959270	16	rs1805005	0.1402	6.317	0.01196
	11	rs3829241	11	rs35264875	2.024	6.314	0.01198
	9	rs683	16	rs1805005	0.216	6.149	0.01315
	14	Rs12896399	16	rs2228479	3.252	6.08	0.01367
	5	rs13289	15	rs1426654	0.5335	5.695	0.01701
	14	rs2402130	15	rs4778241	0.4727	5.647	0.01749
	9	rs683	11	rs1126809	0.3384	5.406	0.02006
	11	Rs1393350	16	rs8051733	0.3785	5.336	0.02088
	16	rs1805007	20	rs6058017	13.9	5.294	0.0214
	15	rs7495174	20	rs1015362	0.3573	5.183	0.02281
	15	rs4778138	15	rs1426654	0.5159	5.138	0.02341
	14	Rs12896399	15	rs4778138	0.4899	5.111	0.02378
	6	rs12203592	11	rs1126809	3.013	5.058	0.02451

	5	rs13289	16	rs8051733	1.889	5.025	0.02498
	11	rs35264875	16	rs1805005	3.627	4.982	0.02562
	15	rs1667394	15	rs1426654	0.5176	4.855	0.02757
	6	rs12203592	9	rs683	0.4829	4.837	0.02785
	6	rs4959270	9	rs10756819	1.849	4.764	0.02906
	11	rs1126809	16	rs1805005	7.679	4.758	0.02916
	6	rs12203592	15	rs1375164	0.4417	4.727	0.02969
	15	rs12592730	20	rs1015362	0.2829	4.725	0.02973
	15	rs1470608	15	rs1426654	0.5145	4.705	0.03007
	6	rs4959270	15	rs1448484	2.69	4.702	0.03012
	15	rs12913832	20	rs6058017	0.428	4.695	0.03026
	6	rs12203592	16	rs2228479	2.329	4.596	0.03206
	12	rs10777129	16	rs2228479	3.246	4.57	0.03253
	5	rs26722	15	rs1470608	12.85	4.529	0.03332
	11	Rs1393350	16	rs1805005	4.493	4.514	0.03362
	6	rs12203592	15	rs4778138	0.3777	4.42	0.03551
	15	rs1375164	15	rs1426654	0.5265	4.385	0.03625
	11	rs3829241	20	rs2378249	0.2916	4.377	0.03642
	15	rs6497292	15	rs1426654	0.5471	4.37	0.03657
	15	rs1667394	16	rs885479	0.477	4.341	0.03721
	5	rs13289	15	rs1545397	0.4738	4.339	0.03725
	9	rs683	16	rs8051733	1.938	4.276	0.03865
	11	rs3829241	14	rs17128291	3.032	4.244	0.03939
	15	rs1800414	16	rs2228479	11.46	4.091	0.04311
Blond/ Non- blond	6	rs12203592	15	rs4778138	4.237	14.86	0.000116
	6	rs12203592	15	rs6497292	6.147	10.54	0.001168
	15	rs4778138	15	rs11636232	5.164	10.05	0.001522
	12	rs12821256	15	rs11636232	0.4045	9.309	0.00228

	16	rs3114908	16	rs1805007	0.3183	8.878	0.002886
	6	rs12203592	15	rs12592730	6.768	8.255	0.004064
	6	rs12203592	15	rs1667394	3.457	7.927	0.00487
	6	rs12203592	15	rs7495174	3.988	7.897	0.004952
	5	Rs16891982	16	rs1805008	16.42	7.858	0.005059
	9	rs683	15	rs1470608	0.3336	7.667	0.005625
	6	rs4959270	16	rs2228479	2.555	7.582	0.005894
	11	rs35264875	15	rs1470608	2.192	7.486	0.006218
	12	rs12821256	15	rs8024968	3.918	7.478	0.006244
	6	rs12203592	15	rs11636232	0.4032	7.357	0.006679
	15	rs1129038	15	rs11636232	3.351	7.079	0.007801
	5	rs13289	9	rs683	0.5515	6.831	0.008961
	9	rs10756819	15	rs1800407	2.635	6.586	0.01028
	9	rs1408799	15	rs4778138	0.2663	6.46	0.01103
	9	rs683	15	rs4778138	0.3042	6.216	0.01266
	5	rs13289	11	rs3829241	1.675	6.201	0.01277
	5	rs13289	6	rs4959270	0.5786	6.197	0.0128
	11	Rs1393350	15	rs4778138	2.198	6.176	0.01295
	12	rs12821256	14	rs2402130	2.427	6.143	0.01319
	11	rs1126809	14	rs17128291	0.3595	6.081	0.01366
	15	rs11636232	15	rs2238289	3.449	6.005	0.01427
	11	rs35264875	14	rs2402130	1.72	5.711	0.01686
	14	rs17128291	15	rs1129038	2.533	5.646	0.01749
	14	rs2402130	15	rs7495174	2.787	5.42	0.01991
	14	rs17128291	16	rs2228479	2.568	5.373	0.02045
	12	rs12821256	15	rs1470608	3.002	5.302	0.0213
	11	rs1126809	16	rs1805007	0.3444	5.168	0.02301
	6	rs12203592	20	rs6119471	12.83	4.905	0.02677

	15	rs1667394	16	rs885479	2.904	4.879	0.02719
	11	rs3829241	15	rs4778241	1.958	4.776	0.02886
	9	rs1408799	15	rs8024968	0.3479	4.755	0.02922
	9	rs1408799	15	rs1470608	0.4087	4.691	0.03032
	9	rs683	15	rs8024968	0.4085	4.681	0.0305
	12	rs10777129	16	rs3114908	0.3492	4.595	0.03206
	14	rs17128291	20	rs6058017	2.42	4.588	0.03219
	11	rs1126809	20	rs6059655	0.363	4.572	0.03251
	6	rs12203592	15	rs1375164	2.324	4.538	0.03316
	15	rs1800407	15	rs1667394	2.882	4.518	0.03353
	12	rs10777129	14	Rs12896399	2.401	4.509	0.03373
	11	rs3829241	15	rs12441727	1.81	4.504	0.03381
	6	rs12203592	11	rs35264875	0.4912	4.489	0.03411
	5	Rs16891982	11	rs1042602	4.887	4.48	0.0343
	11	rs1126809	15	rs7183877	0.1788	4.457	0.03476
	15	rs1545397	16	rs3114908	2.88	4.45	0.03491
	15	rs4778241	15	rs11636232	2.286	4.428	0.03536
	6	rs12203592	15	rs1129038	2.839	4.364	0.03671
	14	rs2402130	20	rs2378249	1.878	4.35	0.03701
	15	rs1470608	20	rs6059655	2.727	4.337	0.0373
	6	rs4959270	9	rs1408799	0.6364	4.321	0.03764
	11	rs3829241	16	rs8051733	1.506	4.303	0.03805
	16	rs2228479	16	rs8051733	0.5072	4.299	0.03813
	15	rs11636232	15	rs1667394	2.747	4.278	0.03861
	9	rs683	15	rs7495174	0.2246	4.265	0.0389
	5	Rs16891982	11	rs35264875	3.438	4.176	0.04101
	5	rs13289	9	rs1408799	0.6248	4.064	0.04381
	11	rs3829241	15	rs4778138	1.935	3.986	0.04587

	14	Rs12896399	15	rs1667394	2.031	3.901	0.04827
	11	Rs1393350	15	rs7495174	2.475	3.9	0.04829
	11	rs35264875	15	rs1545397	2.548	3.855	0.04961
Blond/ Light brown	6	rs12203592	15	rs4778138	4.008	8.367	0.003821
	6	rs12203592	15	rs1667394	5.346	7.97	0.004756
	11	Rs1393350	20	rs6058017	0.276	7.708	0.005498
	6	rs12203592	15	rs6497292	9.91	6.954	0.008364
	15	rs4778241	15	rs2238289	0.07742	6.827	0.00898
	15	rs1129038	16	rs885479	6.397	6.678	0.009761
	11	rs1042602	15	rs4778138	0.3071	6.439	0.01116
	15	rs4778138	15	rs11636232	4.563	6.054	0.01387
	16	rs2228479	16	rs8051733	0.3727	5.971	0.01454
	15	rs1667394	16	rs885479	7.739	5.969	0.01456
	14	rs17128291	15	rs1129038	3.132	5.816	0.01588
	11	rs1126809	15	rs7183877	0.1275	5.577	0.0182
	11	rs35264875	14	rs2402130	1.877	5.515	0.01885
	11	Rs1393350	15	rs7495174	3.778	5.487	0.01916
	15	rs1545397	15	rs4778232	7.706	5.436	0.01973
	6	rs4959270	16	rs2228479	0.4055	5.358	0.02063
	15	rs2238289	16	rs885479	12.92	5.326	0.02101
	11	rs1126809	20	rs6058017	0.2713	5.322	0.02106
	9	rs683	14	rs17128291	0.428	5.295	0.02139
	14	rs17128291	16	rs2228479	2.905	5.172	0.02295
	16	rs3114908	16	rs1805007	0.3525	5.02	0.02505
	14	rs17128291	20	rs6058017	3.217	4.995	0.02542
	11	rs1042602	15	rs1470608	0.3799	4.969	0.02581
	6	rs12203592	15	rs7495174	4.592	4.848	0.02767
	15	rs1129038	16	rs1805008	0.09064	4.839	0.02783

	11	rs1042602	15	rs12441727	0.4832	4.832	0.02794
	15	rs1129038	15	rs11636232	2.976	4.752	0.02926
	16	rs1805007	20	rs1015362	2.778	4.725	0.02973
	6	rs4959270	16	rs1805007	0.3629	4.716	0.02988
	12	rs12821256	15	rs8024968	4.026	4.68	0.03051
	11	rs1126809	16	rs1805007	0.2904	4.656	0.03094
	12	rs10777129	20	rs6059655	5.707	4.604	0.03189
	6	rs4959270	12	rs10777129	0.324	4.583	0.0323
	16	rs8051733	20	rs6059655	2.589	4.506	0.03378
	11	rs1042602	15	rs8024968	0.3853	4.43	0.03532
	14	Rs12896399	20	rs6059655	2.703	4.26	0.03902
	14	rs2402130	15	rs7495174	3.794	4.235	0.03959
	9	rs1408799	15	rs12441727	1.949	4.226	0.0398
	14	rs2402130	15	rs4778138	2.687	4.211	0.04017
	6	rs4959270	15	rs4778232	1.962	4.088	0.04319
	6	rs12203592	15	rs1129038	3.399	4.08	0.0434
	6	rs12203592	15	rs11636232	0.4675	4.065	0.04379
	11	rs1126809	15	rs7495174	4.976	4.016	0.04506
	12	rs12821256	15	rs7183877	4.517	4.006	0.04534
	11	Rs1393350	15	rs4778138	2.037	3.916	0.04784
	9	rs1408799	9	rs683	1.621	3.899	0.04832
	11	rs3829241	15	rs12441727	1.887	3.886	0.04869
	9	rs10756819	20	rs1015362	1.719	3.883	0.04877
	6	rs12203592	16	rs2228479	0.1241	3.882	0.04879
	15	rs1800407	15	rs1667394	3.64	3.852	0.04968
Dark brown/ Non-dark brown	5	Rs16891982	15	rs2238289	0.5261	13.85	0.000199
	12	rs12821256	15	rs11636232	2.527	12.1	0.000503
	5	Rs16891982	15	rs1129038	0.555	9.931	0.001625

	12	rs12821256	16	rs2228479	3.248	9.643	0.001901
	14	Rs12896399	16	rs2228479	0.4319	9.251	0.002353
	5	Rs16891982	15	rs4778241	0.6053	9.122	0.002525
	15	rs6497292	16	rs1805007	5.248	8.898	0.002854
	14	rs17128291	16	rs3114908	0.5143	8.542	0.003471
	15	rs1426654	16	rs885479	0.3837	8.442	0.003667
	5	rs26722	15	rs2238289	0.1614	8.195	0.004201
	5	Rs16891982	16	rs1805005	4.057	8.092	0.004446
	5	Rs16891982	15	rs1667394	0.6245	7.622	0.005765
	11	Rs1393350	15	rs1129038	1.654	7.535	0.00605
	11	rs35264875	16	rs3114908	0.7027	7.475	0.006256
	15	rs1129038	16	rs885479	0.4939	7.416	0.006465
	5	Rs16891982	15	rs6497292	0.5716	7.396	0.006539
	5	rs26722	15	rs1129038	0.3007	7.315	0.00684
	9	rs683	15	rs2238289	0.6485	7.053	0.007915
	5	rs26722	15	rs1470608	0.05722	7.005	0.00813
	15	rs2238289	15	rs1426654	0.5695	6.993	0.008184
	15	rs1129038	15	rs1426654	0.404	6.958	0.008344
	6	rs12203592	15	rs11636232	1.554	6.854	0.008843
	14	Rs12896399	15	rs11636232	0.6521	6.822	0.009005
	6	rs4959270	15	rs11636232	0.6566	6.582	0.0103
	9	rs10756819	11	rs35264875	0.7231	6.52	0.01067
	14	rs17128291	15	rs1375164	1.881	6.502	0.01077
	11	rs1126809	15	rs1129038	1.758	6.411	0.01134
	15	rs1375164	15	rs1667394	0.6716	6.298	0.01209
	11	rs35264875	15	rs11636232	1.372	6.281	0.0122
	15	rs1375164	15	rs4778241	0.6641	6.101	0.01351
	12	rs12821256	20	rs6059655	2.813	5.972	0.01453

	5	rs26722	15	rs1667394	0.2972	5.961	0.01463
	15	rs4778138	16	rs1805007	2.291	5.948	0.01473
	15	rs1129038	16	rs1805007	2.342	5.942	0.01478
	9	rs10756819	15	rs2238289	0.6472	5.91	0.01506
	15	rs1375164	15	rs2238289	0.6689	5.893	0.0152
	16	rs1805005	16	rs8051733	1.615	5.746	0.01653
	14	rs17128291	15	rs1470608	1.916	5.681	0.01715
	11	Rs1393350	15	rs2238289	1.612	5.664	0.01732
	9	rs1408799	11	rs35264875	0.7512	5.635	0.0176
	15	rs2238289	16	rs1805007	2.441	5.635	0.01761
	15	rs1470608	15	rs2238289	0.6645	5.618	0.01778
	15	rs4778241	15	rs1426654	0.5708	5.606	0.0179
	11	rs35264875	20	rs6059655	1.729	5.596	0.018
	9	rs10756819	15	rs1800407	0.5092	5.592	0.01804
	5	Rs16891982	15	rs1800414	0.3496	5.54	0.01858
	11	Rs1393350	15	rs1667394	1.563	5.532	0.01867
	5	Rs16891982	14	Rs12896399	1.52	5.444	0.01964
	9	rs1408799	15	rs2238289	0.6801	5.44	0.01968
	15	rs12592730	16	rs1805007	4.099	5.413	0.01999
	5	Rs16891982	15	rs1470608	0.6829	5.385	0.02031
	15	rs1470608	15	rs1667394	0.6791	5.331	0.02094
	6	rs12203592	9	rs1408799	0.6868	5.303	0.02129
	6	rs12203592	11	Rs1393350	1.475	5.294	0.0214
	11	rs1126809	15	rs6497292	2.312	5.278	0.0216
	5	Rs16891982	15	rs1426654	0.4928	5.276	0.02162
	15	rs4778241	20	rs6059655	2.171	5.244	0.02202
	15	rs1470608	15	rs4778241	0.6642	5.225	0.02226
	11	rs1126809	15	rs12592730	2.657	5.215	0.02239

	9	rs1408799	16	rs1805006	0.2227	5.212	0.02243
	15	rs4778138	15	rs2238289	0.6632	5.104	0.02387
	15	rs7183877	20	rs6119471	0.2082	5.096	0.02398
	6	rs12203592	12	rs12821256	2.037	5.088	0.02409
	15	rs4778241	16	rs885479	0.5005	5.074	0.02429
	15	rs2238289	20	rs6119471	0.4013	5.071	0.02433
	6	rs4959270	9	rs10756819	0.7047	5.069	0.02436
	5	Rs16891982	9	rs683	0.6465	5.049	0.02465
	15	rs1470608	16	rs885479	0.4749	5.034	0.02486
	15	rs1800414	15	rs1426654	0.4135	5.024	0.025
	14	Rs12896399	15	rs6497292	1.741	4.98	0.02564
	15	rs2238289	20	rs6059655	2.129	4.914	0.02664
	11	rs1126809	15	rs1667394	1.724	4.912	0.02666
	15	rs7183877	16	rs8051733	1.667	4.87	0.02732
	15	rs1375164	16	rs885479	0.5153	4.809	0.02831
	15	rs12441727	20	rs1015362	1.638	4.799	0.02848
	9	rs1408799	20	rs6059655	0.5417	4.754	0.02923
	5	rs13289	6	rs4959270	1.4	4.727	0.0297
	15	rs7495174	16	rs1805007	2.64	4.725	0.02972
	5	rs26722	15	rs6497292	0.1878	4.701	0.03015
	9	rs683	15	rs1667394	0.7144	4.689	0.03035
	11	rs1042602	20	rs6059655	0.5158	4.667	0.03074
	5	Rs16891982	15	rs7183877	0.593	4.662	0.03084
	11	rs35264875	20	rs1015362	0.7477	4.635	0.03133
	9	rs1408799	15	rs1667394	0.7156	4.629	0.03144
	11	Rs1393350	15	rs6497292	1.888	4.593	0.0321
	6	rs12203592	12	rs10777129	0.6024	4.581	0.03232
	9	rs683	15	rs1426654	0.5196	4.544	0.03303

	11	rs35264875	16	rs1805007	0.6051	4.531	0.03329
	9	rs10756819	16	rs1110400	10.62	4.519	0.03352
	12	rs10777129	15	rs1448484	1.76	4.478	0.03433
	9	rs683	15	rs1129038	0.7452	4.467	0.03457
	6	rs12203592	20	rs1015362	0.6776	4.463	0.03464
	11	rs35264875	14	rs17128291	1.43	4.453	0.03484
	14	rs17128291	16	rs8051733	0.6486	4.448	0.03495
	15	rs7495174	16	rs885479	0.4518	4.442	0.03506
	5	Rs16891982	15	rs12592730	0.5971	4.429	0.03533
	15	rs8024968	16	rs885479	0.3144	4.317	0.03773
	12	rs12821256	14	rs2402130	0.4883	4.313	0.03782
	6	rs4959270	9	rs1408799	1.344	4.282	0.03851
	15	rs1470608	15	rs4778138	0.6761	4.272	0.03874
	15	rs7495174	20	rs1015362	1.874	4.248	0.03931
	14	rs17128291	16	rs1805005	1.814	4.224	0.03985
	15	rs11636232	20	rs1015362	0.6991	4.178	0.04096
	15	rs11636232	16	rs1805007	0.4996	4.176	0.041
	15	rs1129038	15	rs1667394	0.6779	4.175	0.04104
	14	rs2402130	20	rs2378249	0.6322	4.16	0.04139
	9	rs10756819	15	rs4778241	0.7125	4.129	0.04215
	14	rs2402130	20	rs6058017	0.6016	4.128	0.04219
	11	rs1126809	15	rs2238289	1.689	4.105	0.04275
	11	rs1042602	15	rs1426654	1.838	4.096	0.04299
	11	Rs1393350	11	rs1126809	0.2757	4.09	0.04314
	5	rs26722	15	rs4778138	0.4201	4.07	0.04365
	5	rs26722	15	rs1375164	0.4201	4.069	0.04369
	11	rs35264875	16	rs2228479	1.521	4.007	0.0453
	15	rs1800414	15	rs6497292	0.3311	4.006	0.04533

	15	rs6497292	15	rs1667394	0.5245	3.985	0.04592
	15	rs11636232	16	rs8051733	0.7311	3.955	0.04672
	15	rs1800414	15	rs1667394	0.4517	3.937	0.04724
	9	rs1408799	15	rs1129038	0.7569	3.929	0.04746
	11	rs3829241	20	rs6059655	0.5761	3.922	0.04766
	9	rs1408799	15	rs4778241	0.7303	3.876	0.04899
	11	rs1042602	15	rs7183877	1.689	3.87	0.04916
	5	rs26722	16	rs1805005	9.535	3.844	0.04994
Light brown/ Dark brown	12	rs12821256	15	rs11636232	0.3191	14.3	0.000156
	6	rs12203592	9	rs1408799	1.951	11.2	0.000818
	11	rs35264875	15	rs11636232	0.6359	9.682	0.001861
	9	rs1408799	15	rs1470608	0.4565	9.048	0.00263
	14	rs17128291	16	rs3114908	2.101	8.511	0.00353
	5	Rs16891982	6	rs12203592	3.777	8.211	0.004165
	6	rs12203592	9	rs683	1.786	8.201	0.004186
	9	rs1408799	15	rs1375164	0.5467	8.145	0.004317
	11	rs35264875	15	rs1375164	1.593	7.968	0.00476
	11	rs35264875	15	rs1470608	1.732	7.69	0.005551
	5	Rs16891982	11	Rs1393350	2.547	7.512	0.006129
	12	rs12821256	16	rs2228479	0.2968	7.405	0.006504
	9	rs10756819	11	rs35264875	1.5	7.31	0.006859
	11	rs35264875	14	rs17128291	0.5919	7.256	0.007065
	14	rs17128291	16	rs8051733	1.857	7.214	0.007232
	9	rs683	15	rs1470608	0.5065	7.001	0.008148
	5	rs13289	15	rs1448484	0.285	6.726	0.009504
	9	rs683	15	rs1375164	0.5769	6.693	0.009678
	14	Rs12896399	16	rs2228479	2.252	6.474	0.01095
	12	rs12821256	15	rs1129038	2.414	6.349	0.01174

	5	rs13289	6	rs4959270	0.6309	6.293	0.01212
	6	rs4959270	9	rs1408799	0.6442	6.275	0.01224
	9	rs1408799	15	rs12441727	0.5312	6.183	0.0129
	9	rs1408799	20	rs6059655	2.136	6.164	0.01304
	15	rs12441727	15	rs1129038	0.5003	5.915	0.01501
	11	rs35264875	20	rs6059655	0.5228	5.743	0.01656
	11	rs3829241	15	rs1448484	4.021	5.739	0.01659
	6	rs4959270	15	rs11636232	1.559	5.707	0.0169
	6	rs12203592	9	rs10756819	1.651	5.682	0.01714
	9	rs1408799	14	rs17128291	1.631	5.518	0.01882
	9	rs683	20	rs6059655	2.098	5.484	0.01919
	9	rs1408799	15	rs4778138	0.5681	5.323	0.02104
	16	rs1805005	16	rs8051733	0.5947	5.277	0.02161
	14	rs17128291	16	rs1805007	3.923	5.268	0.02172
	5	rs26722	12	rs12821256	23.95	5.22	0.02233
	6	rs12203592	11	Rs1393350	0.6237	5.204	0.02253
	5	Rs16891982	11	rs1126809	2.802	5.134	0.02346
	9	rs683	14	rs17128291	1.645	5.09	0.02407
	15	rs6497292	16	rs1805008	13.26	5.085	0.02413
	11	rs35264875	15	rs1129038	1.456	5.069	0.02436
	9	rs1408799	11	rs35264875	1.371	5.037	0.02482
	11	rs1042602	15	rs4778138	1.696	4.952	0.02607
	9	rs10756819	15	rs1800407	2.099	4.937	0.02628
	11	Rs1393350	20	rs6058017	1.899	4.914	0.02664
	14	rs17128291	15	rs1375164	0.5251	4.854	0.02758
	15	rs7183877	16	rs8051733	0.5205	4.794	0.02856
	11	rs35264875	16	rs3114908	1.381	4.778	0.02883
	11	rs35264875	20	rs6119471	3.428	4.677	0.03057

	9	rs683	15	rs4778138	0.5979	4.579	0.03236
	15	rs11636232	16	rs1805005	0.5963	4.58	0.03236
	9	rs683	16	rs1805009	28.64	4.566	0.03261
	5	Rs16891982	9	rs1408799	0.4427	4.468	0.03453
	14	rs2402130	20	rs2378249	1.737	4.465	0.03459
	11	rs1042602	15	rs12441727	1.735	4.436	0.03518
	15	rs1470608	15	rs7495174	0.4318	4.368	0.03661
	11	rs1042602	15	rs1470608	1.699	4.358	0.03684
	14	rs17128291	15	rs1470608	0.4935	4.261	0.039
	9	rs683	15	rs11636232	1.439	4.22	0.03995
	9	rs683	11	Rs1393350	1.469	4.2	0.04043
	6	rs12203592	15	rs12441727	1.81	4.162	0.04135
	15	rs8024968	15	rs1426654	3.44	4.138	0.04193
	11	Rs1393350	16	rs1805006	8.352	4.137	0.04196
	15	rs1375164	16	rs1805006	4.632	4.093	0.04307
	9	rs1408799	16	rs1805006	5.44	4.059	0.04393
	5	rs26722	11	Rs1393350	4.996	4.053	0.04409
	6	rs12203592	11	rs1126809	0.578	4.015	0.04509
	14	rs17128291	16	rs1805005	0.5313	3.996	0.04559
	6	rs12203592	14	rs17128291	0.6036	3.961	0.04656
	6	rs4959270	9	rs683	0.7036	3.933	0.04734
	9	rs1408799	15	rs8024968	0.5744	3.933	0.04735
	15	rs4778138	16	rs1805006	14.8	3.914	0.04787
	11	rs1126809	16	rs885479	3.056	3.902	0.04823
	15	rs12592730	16	rs1805008	10.02	3.901	0.04826
	15	rs1129038	16	rs1805005	1.619	3.859	0.04948
Light brown/ Non- light brown	9	rs683	15	rs11636232	1.675	12.05	0.000518
	9	rs683	14	rs17128291	1.944	11.98	0.000538

	9	rs1408799	15	rs1470608	0.4521	11.47	0.000709
	9	rs1408799	15	rs1375164	0.5383	11.35	0.000756
	9	rs683	15	rs1375164	0.5506	10.77	0.001034
	9	rs683	15	rs1129038	0.5922	10.5	0.001192
	12	rs12821256	15	rs11636232	0.4381	10.26	0.001359
	12	rs12821256	15	rs1129038	2.582	10.17	0.001424
	5	Rs16891982	11	Rs1393350	2.482	9.902	0.001651
	11	rs1042602	15	rs4778138	1.938	9.226	0.002386
	9	rs683	15	rs1470608	0.5149	9.076	0.002589
	6	rs12203592	9	rs1408799	1.648	8.577	0.003404
	5	rs26722	12	rs12821256	31.15	8.468	0.003614
	11	rs3829241	15	rs1448484	4.867	8.317	0.003927
	11	Rs1393350	20	rs6058017	2.101	8.214	0.004158
	16	rs2228479	16	rs1805007	4.769	7.779	0.005287
	15	rs1129038	15	rs1667394	0.4814	7.643	0.005701
	5	Rs16891982	11	rs1126809	3.095	7.455	0.006327
	9	rs1408799	15	rs1129038	0.6473	7.209	0.007252
	11	rs1042602	15	rs1470608	1.841	7.185	0.007351
	15	rs6497292	16	rs1805008	6.338	7.153	0.007484
	6	rs12203592	11	rs35264875	0.6722	7.134	0.007562
	5	Rs16891982	6	rs12203592	3.045	7.019	0.008067
	9	rs1408799	14	rs17128291	1.625	6.94	0.00843
	15	rs1129038	16	rs1805008	2.55	6.871	0.00876
	11	rs35264875	15	rs1375164	1.47	6.844	0.008893
	15	rs1448484	15	rs11636232	3.707	6.769	0.009274
	9	rs683	15	rs1667394	0.6091	6.646	0.009941
	9	rs1408799	15	rs12441727	0.5539	6.588	0.01027
	9	rs1408799	15	rs4778138	0.5728	6.547	0.0105

	14	rs17128291	16	rs8051733	1.602	6.528	0.01062
	9	rs683	15	rs4778138	0.5846	6.496	0.01081
	5	rs13289	15	rs1448484	0.33	6.471	0.01097
	9	rs683	11	Rs1393350	1.494	6.466	0.011
	16	rs1805007	16	rs8051733	0.4335	6.356	0.0117
	15	rs1129038	15	rs2238289	0.4688	6.298	0.01209
	11	rs35264875	14	rs17128291	0.6648	6.258	0.01236
	6	rs12203592	9	rs683	1.529	6.251	0.01241
	15	rs4778138	15	rs1129038	0.545	6.188	0.01286
	6	rs4959270	16	rs1805007	0.5018	6.139	0.01323
	6	rs12203592	9	rs10756819	1.554	6.104	0.01348
	11	Rs1393350	16	rs1805006	13.24	5.768	0.01632
	9	rs1408799	15	rs11636232	1.44	5.752	0.01647
	14	rs2402130	14	rs17128291	0.5785	5.665	0.0173
	9	rs1408799	16	rs1805008	2.043	5.53	0.01869
	9	rs683	15	rs4778241	0.6306	5.503	0.01898
	11	rs1042602	16	rs885479	2.429	5.492	0.0191
	11	rs35264875	15	rs11636232	0.7542	5.489	0.01913
	5	Rs16891982	15	rs1129038	0.5325	5.45	0.01957
	9	rs1408799	12	rs12821256	1.835	5.443	0.01965
	5	Rs16891982	9	rs1408799	0.4953	5.428	0.01982
	9	rs10756819	15	rs4778138	0.5905	5.32	0.02108
	15	rs11636232	16	rs1805005	0.6248	5.244	0.02203
	14	rs17128291	16	rs1805007	2.043	5.233	0.02217
	15	rs12592730	16	rs1805008	5.646	5.22	0.02233
	9	rs683	16	rs1805007	1.846	5.199	0.02261
	9	rs1408799	11	Rs1393350	1.417	5.142	0.02335
	5	rs26722	16	rs1805007	14.51	5.142	0.02336

	11	rs1042602	16	rs1805008	0.4833	5.089	0.02408
	16	rs885479	20	rs6059655	4.397	5.034	0.02485
	9	rs1408799	20	rs6059655	1.825	5.017	0.0251
	11	rs1042602	12	rs12821256	0.5293	4.958	0.02596
	5	rs26722	11	Rs1393350	4.847	4.863	0.02743
	5	rs13289	6	rs4959270	0.7081	4.852	0.02762
	5	Rs16891982	15	rs11636232	1.891	4.8	0.02846
	11	rs35264875	15	rs1470608	1.475	4.796	0.02853
	12	rs12821256	15	rs12592730	4.769	4.792	0.0286
	15	rs8024968	15	rs1426654	3.454	4.761	0.02911
	15	rs1375164	16	rs1805008	2.562	4.729	0.02965
	15	rs12441727	15	rs1129038	0.6088	4.713	0.02993
	9	rs1408799	15	rs8024968	0.5818	4.699	0.03019
	15	rs2238289	15	rs1667394	0.4543	4.656	0.03095
	9	rs1408799	15	rs1545397	0.5614	4.645	0.03114
	11	rs3829241	15	rs1545397	1.88	4.643	0.03117
	5	Rs16891982	11	rs3829241	1.972	4.631	0.0314
	9	rs683	16	rs1805008	1.905	4.609	0.03181
	9	rs1408799	16	rs8051733	1.354	4.601	0.03195
	11	rs35264875	20	rs6119471	2.878	4.578	0.03239
	15	rs1470608	16	rs1805008	3.128	4.573	0.03248
	15	rs1470608	15	rs7495174	0.4787	4.565	0.03263
	15	rs1545397	15	rs1800407	3.953	4.529	0.03332
	11	rs1042602	15	rs12441727	1.61	4.523	0.03345
	11	rs3829241	15	rs6497292	2.121	4.518	0.03353
	9	rs10756819	15	rs11636232	1.377	4.488	0.03413
	9	rs683	15	rs8024968	0.6038	4.469	0.03452
	15	rs1375164	15	rs7495174	0.5315	4.449	0.03492

	16	rs3114908	16	rs8051733	0.7368	4.424	0.03543
	9	rs683	20	rs6059655	1.755	4.408	0.03578
	6	rs4959270	12	rs10777129	0.5786	4.362	0.03674
	5	rs13289	15	rs1667394	0.6599	4.354	0.03691
	15	rs1375164	15	rs1129038	0.6892	4.299	0.03813
	9	rs1408799	15	rs4778241	0.6621	4.274	0.0387
	6	rs12203592	14	rs17128291	0.6465	4.241	0.03947
	15	rs4778241	15	rs1129038	0.617	4.225	0.03982
	11	rs3829241	20	rs6119471	4.282	4.208	0.04023
	15	rs1375164	16	rs1805006	4.237	4.126	0.04224
	14	rs17128291	16	rs3114908	1.476	4.079	0.04343
	12	rs12821256	15	rs6497292	3.771	4.067	0.04372
	11	rs1042602	15	rs1375164	1.441	4.006	0.04535
	15	rs1470608	15	rs11636232	1.687	3.995	0.04565
	9	rs1408799	11	rs1126809	1.539	3.986	0.04589
	16	rs1805007	16	rs885479	3.866	3.984	0.04593
	11	rs3829241	15	rs1129038	1.394	3.972	0.04626
	15	rs1129038	15	rs7183877	0.4155	3.934	0.04731
	11	rs1042602	15	rs8024968	1.641	3.919	0.04774
	16	rs1805007	20	rs1015362	0.5609	3.919	0.04775
	9	rs10756819	11	rs35264875	1.287	3.868	0.04923
Red/ Non- red	11	rs3829241	11	rs35264875	0.5357	10.73	0.001055
	9	rs1408799	15	rs11636232	2.016	8.601	0.00336
	16	rs1805008	20	rs2378249	4.102	7.991	0.0047
	15	rs1545397	20	rs6059655	14.37	7.92	0.004889
	9	rs683	14	rs2402130	0.4587	7.439	0.006381
	11	rs35264875	11	rs1126809	0.4652	6.877	0.008733
	11	rs35264875	15	rs1667394	2.075	6.767	0.009287

	6	rs12203592	16	rs1805008	2.847	6.708	0.009599
	16	rs1805005	16	rs2228479	6.805	6.6	0.0102
	14	rs2402130	20	rs6058017	3.199	6.592	0.01025
	9	rs1408799	9	rs683	0.5162	6.493	0.01083
	14	rs17128291	20	rs1015362	2.455	6.351	0.01173
	6	rs12203592	11	Rs1393350	0.5512	6.089	0.0136
	15	rs1800407	15	rs1667394	3.275	5.757	0.01642
	15	rs1448484	16	rs2228479	13.94	5.688	0.01708
	14	Rs12896399	15	rs1470608	2.308	5.613	0.01782
	15	rs4778138	20	rs1015362	0.1741	5.547	0.01851
	9	rs683	11	rs35264875	1.6	5.523	0.01877
	11	rs35264875	16	rs1805007	1.841	5.468	0.01937
	16	rs1805005	16	rs8051733	0.5067	5.302	0.0213
	9	rs10756819	15	rs12441727	2.22	5.238	0.0221
	9	rs10756819	20	rs1015362	0.5267	5.192	0.02269
	15	rs1800407	15	rs2238289	3.152	5.151	0.02324
	9	rs683	12	rs10777129	0.3918	5.125	0.02358
	16	rs1805007	16	rs8051733	2.413	5.067	0.02438
	15	rs6497292	16	rs1805008	9.553	5.008	0.02524
	11	rs1126809	16	rs1805008	0.278	5.002	0.02532
	11	rs1042602	15	rs8024968	2.487	4.981	0.02563
	16	rs1805007	20	rs6059655	0.4301	4.922	0.02652
	11	rs35264875	15	rs6497292	4.041	4.909	0.02672
	11	rs3829241	12	rs10777129	2.239	4.815	0.02821
	5	rs26722	11	rs35264875	11.05	4.706	0.03006
	11	rs1042602	20	rs6059655	2.275	4.695	0.03025
	15	rs11636232	16	rs1110400	0.1525	4.658	0.03091
	15	rs1800407	15	rs7183877	3.485	4.628	0.03145

	12	rs12821256	20	rs6059655	0.2394	4.596	0.03205
	14	rs17128291	15	rs7495174	3.875	4.565	0.03264
	16	rs1805008	20	rs6059655	3.716	4.546	0.03299
	14	Rs12896399	15	rs1375164	1.854	4.504	0.03382
	9	rs683	12	rs12821256	2.238	4.458	0.03473
	12	rs12821256	20	rs2378249	0.2533	4.425	0.03541
	15	rs1800407	15	rs4778241	2.585	4.423	0.03545
	11	rs35264875	16	rs1805005	0.5617	4.387	0.03621
	12	rs12821256	14	Rs12896399	2.197	4.356	0.03688
	15	rs1375164	16	rs1805009	4.224	4.355	0.0369
	5	rs13289	11	rs35264875	1.515	4.332	0.03741
	11	rs35264875	15	rs4778241	1.655	4.319	0.0377
	11	rs35264875	11	Rs1393350	0.6728	4.254	0.03915
	15	rs2238289	16	rs1110400	20.27	4.226	0.03981
	14	Rs12896399	15	rs1667394	1.893	4.195	0.04055
	11	rs35264875	15	rs2238289	1.801	4.191	0.04064
	15	rs2238289	16	rs1805008	3.008	4.185	0.04078
	14	rs17128291	15	rs1448484	5.851	4.176	0.04101
	9	rs1408799	15	rs1800407	2.173	4.149	0.04165
	5	Rs16891982	11	rs35264875	3.159	4.138	0.04193
	16	rs1805007	20	rs2378249	0.4835	4.121	0.04234
	9	rs1408799	20	rs6059655	2.018	4.081	0.04337
	16	rs1805007	16	rs1805008	4.696	4.062	0.04386
	14	Rs12896399	15	rs2238289	1.916	4.021	0.04495
	9	rs1408799	11	rs35264875	1.469	3.986	0.04588
	11	rs35264875	16	rs2228479	0.263	3.947	0.04695
	11	Rs1393350	16	rs1805006	0.1966	3.94	0.04714
	14	rs2402130	16	rs1805005	0.3966	3.91	0.04801

	16	rs2228479	16	rs8051733	0.3369	3.885	0.04873
Red/ Blond	14	rs17128291	20	rs1015362	5.38	8.481	0.003589
	16	rs1805008	20	rs2378249	6.378	7.057	0.007898
	9	rs1408799	9	rs683	0.4143	6.762	0.009313
	11	rs1042602	15	rs8024968	4.367	6.502	0.01078
	6	rs4959270	9	rs10756819	0.4082	6.237	0.01251
	9	rs10756819	20	rs1015362	0.411	6.116	0.0134
	9	rs10756819	15	rs12441727	3.044	5.734	0.01664
	11	rs1126809	16	rs1805008	0.1732	5.621	0.01774
	9	rs10756819	16	rs1805008	0.2249	5.339	0.02085
	5	rs13289	9	rs683	2.148	5.33	0.02097
	16	rs1805005	16	rs2228479	8.864	5.296	0.02138
	16	rs1805005	16	rs8051733	2.505	5.156	0.02317
	11	rs1126809	14	rs17128291	4.18	5.053	0.02458
	12	rs12821256	14	Rs12896399	3.065	5.015	0.02513
	15	rs4778138	20	rs1015362	0.1288	4.979	0.02566
	14	rs17128291	15	rs4778232	4.852	4.968	0.02582
	15	rs11636232	20	rs6058017	0.2001	4.843	0.02775
	9	rs1408799	15	rs4778138	5.419	4.456	0.03479
	11	rs1126809	15	rs7183877	7.817	4.453	0.03483
	15	rs1129038	15	rs11636232	0.2601	4.426	0.03539
	6	rs12203592	20	rs2378249	3.469	4.405	0.03583
	16	rs1805008	20	rs6059655	7.077	4.352	0.03697
	9	rs1408799	15	rs11636232	2.017	4.344	0.03714
	9	rs1408799	20	rs6058017	0.2383	4.327	0.03752
	14	Rs12896399	15	rs12441727	2.694	4.311	0.03786
	12	rs12821256	16	rs1805007	4.016	4.233	0.03966
	6	rs12203592	15	rs7495174	0.1	4.196	0.04051

	11	rs1042602	15	rs4778138	3.437	4.168	0.0412
	9	rs10756819	14	Rs12896399	1.988	4.118	0.04242
	11	rs1042602	15	rs1470608	3.025	4.016	0.04506
	6	rs12203592	15	rs4778138	0.2933	4.011	0.0452
	11	rs1042602	15	rs12441727	2.333	3.931	0.0474
	11	rs35264875	11	rs1126809	0.4652	3.882	0.04881
Red/ Light brown	11	rs3829241	11	rs35264875	0.557	7.629	0.005745
	9	rs10756819	15	rs12441727	3.148	7.471	0.006269
	14	rs2402130	20	rs6058017	8.2	6.922	0.008515
	5	rs13289	15	rs4778138	4.322	5.94	0.0148
	9	rs683	14	rs2402130	0.4582	5.538	0.01861
	16	rs1805008	20	rs2378249	3.481	5.369	0.0205
	15	rs1545397	20	rs6059655	16.99	5.131	0.0235
	16	rs1805007	20	rs2378249	0.4121	5.078	0.02424
	11	rs35264875	16	rs1805007	1.896	5.001	0.02534
	6	rs12203592	16	rs1805008	2.945	4.982	0.02561
	15	rs4778138	16	rs8051733	4.777	4.858	0.02752
	16	rs2228479	16	rs8051733	0.2115	4.817	0.02818
	9	rs683	14	rs17128291	0.4423	4.793	0.02857
	6	rs12203592	9	rs10756819	0.4836	4.749	0.02931
	9	rs683	12	rs10777129	0.3092	4.7	0.03016
	11	rs1042602	16	rs1805008	2.721	4.556	0.03281
	16	rs1805005	16	rs2228479	5.307	4.547	0.03297
	16	rs1805007	20	rs6059655	0.4062	4.437	0.03517
	16	rs1805007	16	rs8051733	2.548	4.329	0.03746
	9	rs10756819	16	rs2228479	0.1057	4.306	0.03799
	15	rs4778138	20	rs1015362	0.184	4.302	0.03806
	14	rs17128291	15	rs7495174	4.65	4.259	0.03904

	14	rs2402130	16	rs1805005	0.3687	4.118	0.04243
	15	rs11636232	16	rs1110400	0.1811	4.094	0.04303
	14	Rs12896399	15	rs1470608	2.196	3.986	0.04589
	14	Rs12896399	15	rs12441727	2.175	3.972	0.04625
	11	rs35264875	15	rs1667394	1.879	3.935	0.04728
	12	rs10777129	16	rs1110400	23.27	3.898	0.04834
	12	rs12821256	14	Rs12896399	2.281	3.855	0.0496

Appendix C Summary of all Interactions identified via MDR and PLINK Analyses

Trait classification	First SNP	Second SNP	MDR-Relief filter (Top 5 and 20 SNPs)	MDR (All SNPs in LD removed)	MDR (without rs12913832)	MDR (without rs16891982)	PLINK
Black/ non-black	rs8051733	rs1129038	-	-	-	0.002	-
	rs12203592	rs4959270	-	-	-	-	0.000006581*
	rs12203592	rs35264875	-	-	-	-	0.00001499*
Black/ Dark brown	rs12203592	rs4959270	-	-	-	-	0.00002356*
Blond/ Light brown	rs12203592	rs12913832	0.007	-	-	-	0.03506
	rs12441727	rs1393350	0.012	-	-	-	-
	rs12203592	rs1129038	0.024	-	0.117	0.117	0.0434
	rs12203592	rs2402130	0.056	0.461	0.566	0.566	-
	rs8051733	rs3829241	-	0.013	0.012	0.012	-
	rs26722	rs2402130	-	0.174	-	-	-
Dark brown/ non-dark brown	rs12203592	rs16891982	0.061	-	-	-	-
	rs12203592	rs28777	-	-	>0.1	-	-

	rs11636232	rs10756819	-	0.026	-	-	-
	rs13289	rs4778138	-	-	-	0.046	-
Red/ non- red	rs35264875	rs1805005	-	0.32	-	-	0.03621
Red/ blond	rs1015362	rs10756819	0.017	-	-	-	0.0134
	rs12896399	rs10756819	0.013	0.001	-	-	0.04242
	rs12896399	rs683	0.019	-	-	-	-
	rs8051733	rs1805005	-	0.129	-	-	0.02317
	rs35264875	rs1805005	-	0.018	-	-	-
Red/ Light brown	rs10756819	rs1805005	-	0.041	-	-	-
	rs12896399	rs10756819	-	0.001	-	-	-
	rs35264875	rs1805005	-	0.046	-	-	-

Appendix D Initial AUC table for all the SNPs and multiplied Diplotypes

Table listing AUC values of 100% training/test runs and 1000 CV runs of HIrisPlex, Multiplied diplotypes (interactions) and SNPs involved in interactions. '100' indicates 'all samples' AUC and 90/10 indicates cross-validated AUCs

		Black (AUC)		Brown(AUC)		Red(AUC)		Blond(AUC)	
Models	**Model type**	**100**	**90/10**	**100**	**90/10**	**100**	**90**	**100**	**90**
HIrisPlex	HirisPlex	0.9473	0.8625	0.8516	0.7616	0.9575	0.9121	0.9012	0.8454
HirisPlex+rs12203592*rs4959270	HirisPlex+ 1 Interaction	0.9489	0.8697	0.8581	0.7678	0.9572	0.9027	0.9018	0.8421
HirisPlex+rs12913832*rs16891982	HirisPlex+ 1 Interaction	0.9462	0.8637	0.8524	0.7594	0.9576	0.9049	0.9023	0.8359
HirisPlex+rs12203592*rs12913832	HirisPlex+ 1 Interaction	0.9474	0.8612	0.8518	0.7541	0.9575	0.8994	0.8998	0.8381
HirisPlex+rs12896399*rs683	HirisPlex+1 interaction	0.9485	0.8626	0.8606	0.7647	0.9606	0.9101	0.9081	0.8425
HirisPlex+rs12203592*rs16891982	HirisPlex+1 interaction	0.9481	0.8652	0.8524	0.7544	0.9575	0.8900	0.9018	0.8402
HirisPlex+rs12203592*rs2402130	HirisPlex+1 interaction	0.9471	0.8591	0.8515	0.7555	0.9565	0.8997	0.9027	0.8403
HirisPlex+rs12203592*rs28777	HirisPlex+1 interaction	0.9472	0.8568	0.8514	0.7520	0.9575	0.8935	0.9017	0.8454
HIrisPlex+rs35264875	HirisPlex+1SNP	0.9469	0.8585	0.8660	0.7714	0.9600	0.9088	0.9121	0.8475
HIrisPlex+rs26722	HirisPlex+1SNP	0.9545	0.8685	0.8584	0.7668	0.9577	0.9101	0.9017	0.8412
HIrisPlex+rs12441727	HirisPlex+1SNP	0.9496	0.8626	0.8546	0.7550	0.9571	0.9083	0.9053	0.8478
HIrisPlex+rs1129038	HirisPlex+1SNP	0.9469	0.8556	0.8543	0.7585	0.9576	0.9100	0.9030	0.8450
HIrisPlex+rs11636232	HirisPlex+1SNP	0.9474	0.8642	0.8528	0.7577	0.9581	0.9077	0.9012	0.8396

HIrisPlex+rs10756819	HirisPlex+1SNP	0.9472	0.8604	0.8537	0.7554	0.9585	0.9092	0.9001	0.8342
HIrisPlex+rs1015362	HirisPlex+1SNP	0.9471	0.8611	0.8524	0.7592	0.9575	0.9055	0.9020	0.8389
HIrisPlex+rs3829241	HirisPlex+1SNP	0.9472	0.8553	0.8516	0.7509	0.9575	0.9052	0.9025	0.8334
HIrisPlex+rs8051733	HirisPlex+1SNP	0.9492	0.8727	0.8492	0.7576	0.9572	0.9033	0.9012	0.8430
HIrisPlex+rs13289	HirisPlex+1SNP	0.9473	0.8590	0.8525	0.7530	0.9569	0.9024	0.8998	0.8335
HIrisPlex+rs4778138	HirisPlex+1SNP	0.9460	0.8603	0.8504	0.7524	0.9594	0.9120	0.8996	0.8343
HirisPlex+rs35264875+rs12203592*rs35264875	HirisPlex+1SNP+1 Interaction	0.9496	0.8624	0.8666	0.7673	0.9629	0.9100	0.9165	0.8498
HirisPlex+rs35264875+rs35264875*rs1805005	HirisPlex+1SNP+1 Interaction	0.9478	0.8544	0.8672	0.7635	0.9609	0.9042	0.9122	0.8486
HirisPlex+rs12441727+rs12441727*rs1393350	HirisPlex+1SNP+1 Interaction	0.9498	0.8385	0.8554	0.7505	0.9643	0.9122	0.9054	0.8438
HirisPlex+rs26722+rs26722*rs2402130	HirisPlex+1SNP+1 Interaction	0.9553	0.8677	0.8588	0.7661	0.9581	0.8955	0.9019	0.8369
HirisPlex+rs8051733+rs8051733*rs1805005	HirisPlex+1SNP+1 interaction	0.9490	0.8566	0.8523	0.7561	0.9604	0.9113	0.9033	0.8380
HirisPlex+rs1129038+rs1129038*rs16891982	HirisPlex+1SNP+1 Interaction	0.9461	0.8604	0.8552	0.7562	0.9582	0.9047	0.9044	0.8475
HirisPlex+rs10756819+rs10756819*rs1805005	HirisPlex+1SNP+1 Interaction	0.9485	0.8601	0.8539	0.7529	0.9598	0.9119	0.9012	0.8331
HirisPlex+rs1129038+rs12203592*rs1129038	HirisPlex+1SNP+1 Interaction	0.9467	0.8539	0.8546	0.7537	0.9583	0.9055	0.9025	0.8354
HirisPlex+rs8051733+rs12203592*rs8051733	HirisPlex+1SNP+1 Interaction	0.9509	0.8667	0.8500	0.7547	0.9584	0.9020	0.9016	0.8364
HirisPlex+rs10756819+rs12896399*rs10756819	HirisPlex+1SNP+1 Interaction	0.9472	0.8556	0.8538	0.7542	0.9585	0.9133	0.9001	0.8304
HirisPlex+rs35264875+rs1129038	HirisPlex+2SNPs	0.9463	0.8570	0.8678	0.7738	0.9607	0.9116	0.9127	0.8499

HirisPlex+rs11636232+rs10756819	HirisPlex+2SNPs	0.9474	0.8555	0.8560	0.7511	0.9591	0.9085	0.8999	0.8333
HirisPlex+rs1015362+rs10756819	HirisPlex+2SNPs	0.9468	0.8552	0.8556	0.7517	0.9589	0.9025	0.9011	0.8333
HirisPlex+rs8051733+rs1129038	HirisPlex+2SNPs	0.9491	0.8601	0.8521	0.7523	0.9576	0.9088	0.9031	0.8407
HirisPlex+rs8051733+rs3829241	HirisPlex+2SNPs	0.9496	0.8602	0.8496	0.7519	0.9575	0.9053	0.9023	0.8361
HirisPlex+rs13289+rs4778138	HirisPlex+2SNPs	0.9460	0.8634	0.8516	0.7554	0.9591	0.9031	0.8997	0.8345
HirisPlex+rs35264875+rs1129038+rs35264875*rs1129038	HirisPlex+2SNPs+1 interaction	0.9466	0.8545	0.8677	0.7697	0.9607	0.9114	0.9129	0.8504
HirisPlex+rs1015362+rs10756819+rs1015362*rs10756819	HirisPlex+2SNPs+1 interaction	0.9456	0.8539	0.8600	0.7572	0.9651	0.9126	0.9039	0.8369
HirisPlex+rs11636232+rs10756819+rs11636232*rs10756819	HirisPlex+2SNPs+1 interaction	0.9474	0.8501	0.8603	0.7549	0.9609	0.9108	0.9019	0.8349
HirisPlex+rs8051733+rs3829241+rs8051733*rs3829241	HirisPlex+2SNPs+1 interaction	0.9504	0.8569	0.8555	0.7565	0.9575	0.9028	0.9056	0.8373
HirisPlex+rs8051733+rs1129038+rs8051733*rs1129038	HirisPlex+2SNPs+1 interaction	0.9489	0.8525	0.8549	0.7482	0.9576	0.9015	0.9065	0.8360
HirisPlex+rs13289+rs4778138+rs13289*rs4778138	HirisPlex+2SNPs+1 interaction	0.9449	0.8530	0.8550	0.7530	0.9620	0.9058	0.9023	0.8325

Appendix E Pearson Correlation Matrix

Pearson Correlation matrix for Interactions and the individual markers in the interactions with the corresponding p-value. The values were generated using IBM© SPSS statistics software. Highlighted in pink are the SNPs or interactions for which *p*-value < 0.05.

SNPs	Black	Black (p)	Brown	Brown (p)	Red	Red (p)	Blond	Blond (p)
N29insA_A	-0.014	0.673	-0.019	0.570	0.065	0.057	-0.019	0.5791754
rs11547464_A	-0.002	0.950	-0.060	0.079	.086*	0.011	0.007	0.8267255
rs885479_A	.180**	0.000	-0.028	0.409	-0.055	0.104	-0.053	0.1178496
rs1805008_T	-.076*	0.026	-.082*	0.016	.103**	0.002	.082*	0.0154798
rs1805005_T	-.110**	0.001	.074*	0.028	0.000	0.993	-0.013	0.6927419
rs1805006_A	-0.028	0.415	0.031	0.355	0.032	0.341	-0.049	0.148846
rs1805007_T	-.098**	0.004	-.190**	0.000	.378**	0.000	0.002	0.9536193
rs1805009_C	-0.027	0.433	-.096**	0.005	.162**	0.000	0.010	0.763966
Y152OCH_A	-0.010	0.766	-0.050	0.139	.102**	0.002	-0.013	0.695124
rs2228479_A	0.007	0.836	-0.002	0.958	-.083*	0.014	.070*	0.0401314
rs1110400_C	-0.043	0.202	0.011	0.740	0.060	0.077	-0.033	0.3271915
rs28777_C	.493**	0.000	-.125**	0.000	-.101**	0.003	-.137**	4.745E-05
Rs16891982_C	.544**	0.000	-.135**	0.000	-.112**	0.001	-.144**	2.428E-05
rs12821256_C	-.105**	0.002	-.072*	0.034	0.050	0.146	.136**	6.174E-05
rs4959270_C	.096**	0.005	-.096**	0.005	-0.025	0.459	.077*	0.0240691
rs12203592_T	-0.047	0.171	.171**	0.000	0.012	0.716	-.205**	1.275E-09
rs1042602_A	-.101**	0.003	-0.010	0.765	0.013	0.698	.081*	0.017495

rs1800407_A	-.095**	0.005	0.039	0.255	0.051	0.136	-0.021	0.5368742
rs2402130_G	.081*	0.017	-0.045	0.185	0.038	0.267	-0.037	0.2747049
rs12913832_T	.387**	0.000	0.024	0.475	-.110**	0.001	-.239**	1.426E-12
rs2378249_G	-0.025	0.469	-0.005	0.891	0.010	0.764	0.017	0.6121742
Rs12896399_T	-.095**	0.005	0.009	0.801	-0.053	0.116	.113**	0.0008683
Rs1393350_A	-.147**	0.000	-0.002	0.960	.121**	0.000	0.011	0.7478788
rs683_C	.288**	0.000	-0.044	0.192	-0.059	0.082	-.120**	0.0004024
rs8051733_C	0.044	0.208	-.224**	0.000	.196**	0.000	.097**	0.0052735
rs1129038_G	.374**	0.000	0.030	0.384	-.110**	0.002	-.238**	4.777E-12
rs12441727_A	0.068	0.056	-0.042	0.232	-0.036	0.309	0.036	0.3038206
rs26722_T	.265**	0.000	-.096*	0.012	-0.019	0.624	-0.070	0.0693098
rs35264875_T	.117**	0.001	-.170**	0.000	-0.027	0.428	.162**	1.775E-06
rs11636232_T	-.239**	0.000	-0.036	0.296	.102**	0.003	.153**	7.637E-06
rs13289_G	.120**	0.000	-0.052	0.128	-0.002	0.944	-0.024	0.4867847
rs1015362_T	0.013	0.696	0.008	0.822	0.033	0.341	-0.050	0.1458979
rs10756819_G	.137**	0.000	-0.032	0.359	-0.044	0.204	-0.025	0.4644885
rs3829241_A	-.139**	0.000	-0.014	0.690	0.065	0.058	.071*	0.0381154
rs4778138_G	.253**	0.000	0.002	0.955	-.097**	0.004	-.115**	0.0007829
rs8051733X1129038	.296**	0.000	-0.052	0.136	-0.011	0.746	-.153**	1.185E-05
rs12203592X4959270	-.116**	0.001	.135**	0.000	0.027	0.423	-.116**	0.0006602
rs12203592X35264875	-0.060	0.080	.098**	0.004	-0.013	0.709	-.074*	0.0302863
rs1129038X16891982	.572**	0.000	-.153**	0.000	-.101**	0.004	-.137**	9.23E-05
rs12913832X16891982	.573**	0.000	-.156**	0.000	-.106**	0.002	-.143**	2.514E-05
rs12203592X12913832	0.054	0.115	.096**	0.005	-0.043	0.211	-.134**	8.189E-05

rs12441727X1393350	-0.043	0.229	-0.055	0.126	0.048	0.177	0.064	0.0741338
rs12203592X1129038	0.059	0.093	.092**	0.009	-0.038	0.282	-.136**	0.0001034
rs12203592X2402130	0.019	0.586	0.045	0.184	0.039	0.253	-.110**	0.0012264
rs8051733X3829241	-.103**	0.003	-.156**	0.000	.153**	0.000	.156**	8.234E-06
rs26722X2402130	.154**	0.000	-0.034	0.374	-0.027	0.481	-0.057	0.1386621
rs35264875X1129038	.383**	0.000	-0.026	0.454	-.112**	0.001	-.158**	6.817E-06
rs12203592X16891982	0.060	0.080	0.021	0.544	-0.033	0.330	-0.046	0.1814014
rs12203592X28777	.078*	0.023	0.002	0.949	-0.027	0.422	-0.040	0.2473617
rs11636232X10756819	-.117**	0.001	-0.007	0.852	0.032	0.367	.074*	0.0364618
rs13289X4778138	.267**	0.000	-0.067	0.053	-0.050	0.145	-.074*	0.0316865
rs12203592X8051733	-0.011	0.744	0.011	0.745	.125**	0.000	-.118**	0.00079
rs35264875X1805005	-.093**	0.007	-0.002	0.951	0.033	0.331	0.047	0.1734171
rs1015362X10756819	.079*	0.025	-0.002	0.944	-0.054	0.125	-0.012	0.7297947
rs12896399X10756819	0.010	0.783	-0.016	0.639	-0.021	0.557	0.033	0.3429666
rs12896399X683	0.056	0.098	-0.010	0.760	-0.036	0.285	0.001	0.9828077
rs8051733X1805005	-.078*	0.024	-0.016	0.638	0.020	0.576	0.068	0.0525139
rs10756819X1805005	-0.065	0.061	0.053	0.129	0.021	0.544	-0.039	0.2619905
rs35264875X12913832	.401**	0.000	-0.032	0.351	-.119**	0.001	-.166**	1.101E-06

Appendix F Sequential Model Building AUC Summary

Highlighted in yellow is the final model that could be added on to HIrisPlex

Models	Model #	Black 100	Black 90	Brown 100	Brown 90	Red 100	Red 90	Blond 100	Blond 90
HirisPlex	1	0.9473	0.8625	0.8516	0.7616	0.9575	0.9121	0.9012	0.8454
HIrisPlex+rs12896399X683+rs35264875_T+rs12203592X35264875	2	0.9508	0.8696	0.8698	0.7728	0.9648	0.9116	0.9160	0.8501
HIrisPlex++rs12896399X683+rs35264875_T+rs12203592X35264875+rs26722_T	3	0.9583	0.8748	0.8770	0.7769	0.9653	0.9147	0.9159	0.8485
HirisPlex+rs12896399X683+rs35264875_T+rs12203592X35264875+rs26722_T+rs1015362_T+rs10756819_G+rs1015362X10756819	4	0.9618	0.8693	0.8878	0.7780	0.9725	0.9209	0.9219	0.8461
HirisPlex++rs12896399X683+rs35264875_T+rs12203592X35264875+rs26722_T+rs1015362_T+rs10756819_G+rs1015362X10756819+rs12441727_A+rs12441727X1393350	5	0.9652	0.8430	0.8916	0.7662	0.9759	0.9208	0.9217	0.8366

Appendix G Distribution of 3 Identified Interactions and 1 Independent SNP with Hair Color Phenotypes

	rs12896399*rs683				rs12203592*rs35264875				rs1015362*rs10756819				rs26722		
Hair Color	**"0" (%)**	**"1" (%)**	**"2" (%)**	**"4" (%)**	**"0" (%)**	**"1" (%)**	**"2" (%)**	**"4" (%)**	**"0" (%)**	**"1" (%)**	**"2" (%)**	**"4" (%)**	**"0" (%)**	**"1" (%)**	**"2" (%)**
Black (71)	43 (60.6)	8 (11.3)	15 (21.1)	5 (7)	58 (81.7)	0 (0)	8 (11.3)	0 (0)	44 (62)	8 (11.3)	10 (14.1)	3 (4.2)	37 (52.1)	14 (19.7)	1 (1.4)
Brown (599)	357 (59.6)	132 (22)	90 (15)	16 (2.7)	428 (71.5)	58 (9.7)	98 (16.4)	4 (0.7)	406 (67.8)	94 (15.7)	47 (7.8)	9 (1.5)	466 (77.8)	20 (3.3)	2 (0.3)
Red (86)	54 (62.8)	22 (25.6)	8 (9.3)	2 (2.3)	68 (79.1)	4 (4.7)	12 (14)	1 (1.2)	67 (77.9)	8 (9.3)	3 (3.5)	2 (2.3)	58 (67.4)	3 (3.5)	0 (0)
Blond (116)	72 (62.1)	15 (12.9)	25 (21.6)	4 (3.4)	101 (87.1)	2 (1.7)	12 (10.3)	1 (0.9)	77 (66.4)	23 (19.8)	7 (6)	1 (0.9)	75 (64.7)	1 (0.9)	0 (0)