

Victoria University
School of Biomedical and Health Sciences

Blonde Hair Colour:
Classification, Characterisation and Genetic Associations
for use in Forensic Science

By
Michelle Vaughn

A thesis submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy.

April 2010

Doctor of Philosophy Declaration

“I, Michelle Vaughn, declare that the PhD thesis entitled *Blonde Hair Colour: Classification, Characterisation and Genetic Associations for use in Forensic Science* is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work”.

Signature:



Date:

March 24, 2010

This work is dedicated to my Grandparents

Charles and Robina Vaughn
&
Samuel James and Cecile Zacharias

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Swati Baindur-Hudson and Roland van Oorschot for their excellent supervision over the last four years. I thank you for your tremendous support in all stages of this project.

I am very grateful for the support and love I have received from my family and friends in Canada, and most importantly, thanks to Mom and Dad for instilling the thirst for knowledge in me and the life lessons that have gotten me this far.

My gratitude goes to those who have helped me with various aspects of this research program. A special thanks goes to Elizabeth Brooks, Bruce Comber and James Robertson from the Australian Federal Police, Jessica Chang from Victoria Police Forensic Services Department, Domenico Caridi from Victoria University and Angela van Daal from Bond University. I would also like to thank the staff at Vic Uni who make everything possible – Dale Tomlinson, Joe Pelle, Stacey Loyd, Ira Prasatya, Min and Leslie Birch.

My wonderful Uni friends deserve my thanks for their support and friendship – Shruti, Sudinna, Debbie, Anne Lise, Warren, Wun, Lata, Aida, Lydia, Danusha, Julia, Roshani and Gavin.

A very special thanks goes to my Aussie friends Quil, Callie and Kash. I will always value your friendship and will remember the great times we have had for ever. I would also like to thank Jeff and Diana Bryar for accepting me like a member of the family and helping to make my four years in Australia a fantastic experience.

ABSTRACT

The vast variety of normal pigmentation variation is a huge genetic puzzle and hair colour is one part of this puzzle. It is hypothesised that there are genetic variants specifically associated with an objectively defined Blonde phenotype. The first objective of this study was to examine the natural, adult hair colour variation among a population of European ancestry and to evaluate methods for both the measurement and classification of hair colours. The second objective was to characterise the microscopic and chemical features of hair colour in this sample and to determine how they relate to the macroscopic colour. The final objective was to compare genetic results with objectively measured and defined hair colour phenotypes in the search for genetic associations with 'blonde' hair colour. Having this information will assist in developing forensic predictive tests for physical features.

The presented data show that self-reported hair colours and observer-reported colours are similar; however, these categories are not necessarily the best way to categorize hair colour for quantitative research. Using a two-step cluster analysis, hair colour can be divided into categories or clusters based on spectrophotometric measurements in the CIE $L^*a^*b^*$ colour space and these clusters can be well discriminated from each other.

Being able to measure hair colour using digital image analysis may also be of interest. Using the CIE $L^*a^*b^*$ colour space, the hair colour of 196 individuals of European ancestry was measured by both reflective spectrophotometry and by digital image analysis (in V++). Only moderate correlations were found and it is concluded that, although more convenient, hair colour measurement from digital images has limited use in genetics or forensic science when accurate and consistent measurements are required, but may be useful in other less stringent fields such as intelligence or cosmetics.

Microscopic images of hair from 36 individuals were evaluated for colour (using the CIE $L^*a^*b^*$ colour space), hair thickness and the pattern of pigmentation. These measurements were compared to each other and to reflective spectrophotometric (RS) measurements of macroscopic hair colour. Despite a large ΔE^* value, there is a strong relationship between macroscopic and microscopic L^* values that would be useful if one was trying to predict the colour of a head of hair (macroscopic) from a strand (microscopic).

Eumelanin is a black/brown biopolymer pigment that is a major contributor to hair colour, along with other pigments. The eumelanin pigment component in hair was evaluated by quantifying the degradation product pyrrole-2,3,5-tricarboxylic acid (PTCA) by HPLC-UV. The amount of PTCA in the hair of 58 individuals was compared to macroscopic colour measurements using reflective spectrophotometric (RS) and microscopic colour measurements if available. There appears to be a significant relationship between several of the macroscopic and microscopic colour components and the amount of PTCA quantified, such as with the macroscopic RS b^* value and with the microscopic L^* .

Tests for genetic associations were carried out on both categorical variables (clusters) and quantitative variables (L^* , a^* and b^* colour values). In the categorical analysis, five of 58 SNPs show various significant associations with Fair/Blonde hair. These SNPs are in the genes MC1R (R151C), DTNBP1 (H297Y), TYR (R402Q), ATP7B (R832K) and MYO7A (C1628S). MC1R (R151C) and TYR (R402Q) were found to be associated with Red hair. Associations were also found with Dark hair in this sample; a SNP in the gene MATP (L374F) was significantly associated with dark hair in both the allelic and dominant model. Two SNPs in ATP7B (S406A and V456L) were also found to be associated in a dominant model with Dark hair.

In the quantitative variable analysis, polymorphisms in the genes TYR, MATP, DTNBP1, ATRN, MC1R, MLPH, MYO7A and MYO5A were found to be associated with L^* values and these results tend to correspond with genes that were associated with light or dark hair in the categorical analysis.

Future directions of this research could include similar studies with increased sample sizes, greater investigation into the cell biology of hair and the genotyping of more SNPs.

In conclusion, this study successfully met its aims of objectively classifying and characterising blonde (or fair) hair colour and finding genetic associations with blonde hair colour. The results obtained in this study will benefit future research into the genetics of hair colour for use in forensic science and possibly in other areas of pigmentation and forensic research.

PUBLICATIONS

Peer-Reviewed Articles:

Vaughn M., Brooks, E., van Oorschot, R.A.H., Baidur-Hudson, S. **A comparison of macroscopic and microscopic hair colour measurements and a quantification of the relationship between hair colour and thickness.** *Microscopy and Microanalysis* (2009) Volume 15 Pages 189-193

Vaughn, M., van Oorschot, R.A.H., Baidur-Hudson, S. **A comparison of hair colour measurement by digital image analysis with reflective spectrophotometry.** *Forensic Science International* (2009) Volume 183 Pages 97-101

Vaughn, M., van Oorschot, R., Baidur-Hudson, S. **Hair color variation and measurement.** *American Journal of Physical Anthropology* (2008) Volume 137 Issue 1 pp. 91-96

Conference Presentations (Oral):

10/08 ANZFSS 19th International Symposium on the Forensic Sciences: Melbourne, VIC

Michelle Vaughn, Roland van Oorschot and Swati Baidur-Hudson
Blonde hair colour: an overview of the characterisation and genetic associations

Conference Presentations (Posters):

10/08 ANZFSS 19th International Symposium on the Forensic Sciences: Melbourne, VIC

Michelle Vaughn, Roland van Oorschot and Swati Baidur-Hudson
Hair colour measurement by reflective spectrophotometry and digital image analysis

Michelle Vaughn, Elizabeth Brooks, Bruce Comber, Roland van Oorschot and Swati Baidur-Hudson
Characterisation of microscopic hair colour and texture and a comparison with macroscopic colour

Michelle Vaughn, Domenico Caridi, Roland van Oorschot and Swati Baidur-Hudson
Eumelanin characterisation in hair compared to colour measurements

Michelle Vaughn, Angela van Daal, Roland van Oorschot and Swati Baidur-Hudson
Preliminary investigation into the genetic associations with blonde hair colour **Best Biology Poster Award Winner**

04/06 ANZFSS 18th International Symposium on the Forensic Sciences: Fremantle, WA

M. Vaughn, R. van Oorschot and S. Baidur-Hudson
Hair colour measurement and classification

ABBREVIATIONS

2-S-CD	2-S-Cystenyldopa
3-AHP	3-amino-4-hydroxyphenylalanine
3-AT	3-aminotyrosine
DHI	5,6-dihydroxyindole
DHICA	5,6-dihydroxyindole-2-carboxylic acid
5-S-CD	5-S-Cystenyldopa
AIM(s)	ancestry informative marker(s)
BGA	biogeographical ancestry
ΔE^*	Change in Colour
ΔL^*	Change in Lightness
CCD	charge coupled device
CODIS	Combined DNA Index System
CIE	Commission Internationale de l'Eclairage
Co	cortex
Cu	cuticle
CYM	Cyan-Yellow-Magenta
DNA	deoxyribonucleic acid
ddH ₂ O	distilled and deionised water
DI	digital image
E#	Euclidian Cluster Number
FBI	Federal Bureau of Investigation
FD	fluorescence detection
g	gravities
HWE	Hardy-Weinberg Equilibrium
HPLC	high performance liquid chromatography
HSV	Hue, Saturation, Value
LD	linkage disequilibrium
LC/MS	liquid chromatography/mass spectrometry
LL#	Log-likelihood Cluster Number
mAU	milli Active Unit
Md	medulla

MSHR	melanocyte stimulating hormone receptor
MF	microfibril
μL	microlitre
μm	micrometre
μmol	micromolar
Mc	microscopic
mL	millilitre
mm	millimetre
NCIDD	National Criminal Investigation DNA Databank
nm	nanometre
ORC	observer-reported colour
OR	odds ratio
ppb	parts per billion
ppm	parts per million
PCR	polymerase chain reaction
PTCA	pyrolle-2,3,5-tricarboxylic acid
PDCA	pyrolle-2,3-dicarboxylic acid
qPCR	quantitative polymerase chain reaction
RGB	Red-Green-Blue
RS	reflective spectrophotometry
SRC	self-reported colour
STR	short tandem repeat
SNP(s)	single nucleotide polymorphism(s)
4-AHP	4-amino-3-hydroxyphenylalanine
U	Unit
UV	ultraviolet

TABLE OF CONTENTS

Title Page	i
Declaration	ii
Dedication	iii
Acknowledgements	iv
Abstract	v
Publications	vii
Abbreviations	viii
Table of Contents	x
List of Figures	xiii
List of Tables	xv
Chapter 1 Introduction.....	1
1.1 General Introduction.....	1
1.1.1 Introduction	1
1.1.2 Research Outline and Objectives	2
Chapter 2 Background and Literature Review.....	5
2.1 Forensic Biology.....	5
2.1.1 Forensic Hair Examination	5
2.1.2 DNA.....	9
2.1.3 Future Directions in Forensic Science	11
2.2 Phenotype Profiling	12
2.2.1 Ancestry	12
2.2.2 Indirect Phenotype Inference.....	14
2.2.3 Direct Phenotype Inference	14
2.3 Direct Inference of Hair Colour.....	15
2.3.1 Human Hair Colour.....	15
2.3.2 Genetic Studies of Hair Colour	17
2.3.3 Issues in the Study of Human Hair Colour Genetics	18
2.3.4 Continuing the Study of Hair Colour	20
Chapter 3 Hair Colour Measurement and Variation.....	21
3.1 Introduction.....	21
3.1.1 Colour and Perception.....	21
3.1.2 Colour Measurement.....	22
3.1.3 Aims	26
3.2 Materials and Methods.....	27
3.2.1 Subjects.....	27
3.2.2 Sampling	27
3.2.3 Statistics	29
3.3 Results	30
3.3.1 Reported Colours.....	30
3.3.2 Population Variation	30
3.3.3 Grouping the Population	31
3.3.4 Comparing Clusters to the Observer-Reported Colours	35
3.4 Discussion	37

Chapter 4	A comparison of Hair Colour Measurement by Digital Image Analysis with Reflective Spectrophotometry.....	39
4.1	Introduction.....	39
4.1.1	Digital Photography and Digital Colour Measurement	39
4.1.2	Aims	42
4.2	Materials and Methods.....	43
4.2.1	Subjects.....	43
4.2.2	Sampling	43
4.2.3	Digital Image Preparation and Analysis	44
4.2.4	Correction	45
4.2.5	Program Comparison	45
4.2.6	Statistics	46
4.3	Results	47
4.3.1	Colour Correction.....	47
4.3.2	Program Comparison	47
4.3.3	Intra-individual Comparison.....	48
4.3.4	Digital Image Comparison to Reflective Spectrophotometry	48
4.4	Discussion	54
Chapter 5	Microscopic Examination of Hair Colour and Colour Morphology	56
5.1	Introduction.....	56
5.1.1	Microscopic Hair Structure	56
5.1.2	Cellular Biology of Pigmentation	57
5.1.3	Microscopic Hair Colour Measurement	58
5.1.4	Aims	59
5.2	Materials and Methods.....	60
5.2.1	Subjects and Sampling.....	60
5.2.2	Preparation of Hairs for Microscopic Analysis	60
5.2.3	Generation and Collection of Hair Colour Data.....	61
Equipment and Software	61
5.2.4	Generation and Collection of Hair Thickness Data	65
5.2.5	Pigmentation Pattern	65
5.3	Results	66
5.3.1	A Comparison of Microscopic Colour to Reflective Spectrophotometry	66
5.3.2	A Comparison of Microscopic Colour to Macroscopic Digital Image Analysis	67
5.3.3	A Comparison of Microscopic Colours to Reflective Spectrophotometry Clusters	67
5.3.4	A Comparison of Microscopic Colours to Hair Thickness	67
5.3.5	A Comparison of Macroscopic Colour to Hair Thickness	67
5.3.6	Pigment Pattern Analysis.....	68
5.4	Discussion	70
Chapter 6	Chemical Analysis of Hair Pigments	72
6.1	Introduction.....	72
6.1.1	Chemistry.....	72
6.1.2	Hair Pigment Analysis; Methods, Uses and Issues	75
6.1.3	Aims	77
6.2	Materials and Methods.....	78
6.2.1	Sampling	78
6.2.2	Chemicals and Buffers Used in Eumelanin Analysis	78
6.2.3	Hydrogen Peroxide Degradation of Hair for Eumelanin Analysis	78
6.2.4	HPLC Conditions for Eumelanin (PTCA) Analysis	79
6.2.5	Chemicals and Buffers Used in Pheomelanin Analysis	79

6.2.6	LC/MS Conditions for Pheomelanin Analysis	80
6.2.7	HPLC Conditions for Pheomelanin using Fluorescent Detection.....	80
6.3	Results	82
6.3.1	Method Validation for Eumelanin (PTCA) Analysis	82
6.3.2	PTCA content of Hair Samples	87
6.3.3	A Comparison of PTCA Content to Colour Measured by Reflective Spectrophotometry.....	88
6.3.4	A Comparison of PTCA Content to the Microscopic Colour	90
6.3.5	A Comparison of PTCA Content to Hair Thickness.....	91
6.3.6	PTCA Content of Reflective Spectrophotometry Clusters.....	92
6.4	Discussion	93
Chapter 7	Genetic Analysis.....	95
7.1	Introduction.....	95
7.1.1	Genetic Control of Pigmentation.....	95
7.1.2	Genetic Association Studies of Hair Colour.....	98
7.1.3	Aims	98
7.2	Materials and Methods.....	99
7.2.1	Sampling	99
7.2.2	Single Nucleotide Polymorphisms	99
7.2.3	DNA Extraction, Quantification, Preparation and Genotyping	101
7.2.4	Software and Statistical Analysis	102
7.3	Results	103
7.3.1	Polymorphic Single Nucleotide Polymorphisms	103
7.3.2	Hardy-Weinberg Equilibrium and Linkage Disequilibrium.....	104
7.3.3	Tests for Associations with Categorical Variables (Clusters).....	106
7.3.4	Tests for Associations with Quantitative Variables	110
7.3.5	Genetic Interactions	118
7.4	Discussion	120
7.4.1	Associations with Categorical Variables (Clusters).....	120
7.4.2	Associations with Quantitative Variables (L^* , a^* and b^*).....	122
7.4.3	Genetic Interactions	123
7.4.4	Conclusions	123
Chapter 8	Discussion and Conclusions	125
8.1	Summary	125
8.2	Discussion	127
8.3	Future Considerations	132
8.4	Conclusion	134
Chapter 9	References.....	135
Appendix 1 – Raw Data (RS, DI, micro, chem.)		145
Appendix 2 – Computer Code		153
Appendix 3 – Project Information Sheet and Consent Forms		158
Appendix 4 – Published papers		165
Appendix 5 – Genotype Data		CD
Appendix 6 – DI Photographs with white balances		CD
Appendix 7 – microscopy photos, extracted microscopy photos.....		CD

LIST OF FIGURES

Figure 2.1	Hair examination checklists for the Australian Federal Police	7
Figure 2.2	A nine-locus (+ amelogenin) profile of a female that may be used for identification..	11
Figure 2.3	A hypothetical pedigree of one individual who has 75% European and 25% African genealogical ancestry, but approximately 61% European and 39% African genomic ancestry due to random assortment of chromosomes in meiosis. This example does not take into account any recombination of chromosomes.	13
Figure 2.4	Graphical representation of a melanocyte showing some important proteins involved in melanocyte development and the pigmentation pathway (as thought to occur in the epidermis), including the following: 1) the tyrosinase enzyme complex (TYR, TRP1, DCT) located on the membrane of the melanosome, which is responsible for the enzymatic conversion of the amino acid tyrosine into melanin; 2) other proteins located within the melanosomes that play a critical role in melanogenesis (MATP, P gene, SILV, SLC24A5); 3) the signalling pathways affecting the regulation of melanin synthesis, including hormones and receptors (a-MSH, MC1R, ASIP, ATRN); 4) the transcription factors involved in melanin production (PAX3, MITF, SOX10); 5) proteins involved in melanosome transport (MYO5A, MYO7A, RAB27A) and melanosome construction/protein routing (CHS1, HPS1-6); and 6) developmentally important ligands (EDN3, KITLG) and receptors (KIT, EDNRB) which control melanoblast migration and differentiation [34].	16
Figure 3.1	Energy spectrum including the wavelengths of visible light that appear as colours ..	21
Figure 3.2	The HSV colour space showing hue (H), saturation (S) and brightness value (V)	22
Figure 3.3	Diagram of the Red-Green-Blue (RGB) colour model.	23
Figure 3.4	Diagram of the Cyan-Magenta-Yellow (CMY) colour model.	23
Figure 3.5	An illustration of the CIE L*a*b colour space showing the three axes: L* (Black = 0 to White = 100), a* (-100 = green to +100 = red) and b* (-100 = blue to +100 = yellow).[57].	24
Figure 3.6	An example of a 5 unit difference along the L*, a* and b* axis.....	25
Figure 3.7	The locations on the head where reflective spectrophotometry measurements were taken; 5 measurements were taken at each site.	28
Figure 3.8	Population Sample; the non-European samples (NE) and the mature-white haired samples (W) are circled. The scales of the CIE L*a*b* colour space are; L* = 0(black) to 100 (white), a* = -100 (green) to +100 (red) and b* = -100 (blue) to +100 (yellow). Only the relevant region of the scale is shown.....	31
Figure 3.9	Discriminant scores for each number of clusters and the method by which they were determined.	32
Figure 3.10	Five clusters determined by the log-likelihood probability method (LL5) separated according to the results of the first two significant discriminant functions. $F1 = -7.359 + 0.089L^* + 0.953a^* + 0.518b^*$ and $F2 = -9.759 + 0.373L^* - 0.822a^* - 0.233b^*$	34
Figure 3.11	The two most important functions used to discriminate the six clusters as determined by log-likelihood (LL6) were applied to the L*a*b* data and grouped by Cluster (A) and by observer-reported colour (B). $F1 = -8.558 + 0.101L^* + 0.972a^* + 0.435 b^*$ and $F2 = -9.671 + 0.376L^* - 0.908a^* - 0.222b^*$	36
Figure 4.1	Image formation by silver halide and CCD sensors [62]	39
Figure 4.2	Colours and values within the RGB model	40
Figure 4.3	Sample photographs before cropping and analysis	43
Figure 4.4	The relationship between RS and DI measurements for i) L* ii) a* and iii) b*	50
Figure 4.5	RS Clusters with RS data (A) compared to DI data (using RS discriminant functions) (B) for LL5. Clusters are numbered 1 to 5.	53
Figure 5.1	A cross section of a hair follicle.	56
Figure 5.2	Illustrates the microscopic features of the hair shaft. a) Cartoon of the hair shaft with cut-through to reveal inner aspects. Co, cortex; Cu, cuticle; Md, medulla. A microfibril (300 nm) of single cortical cell is revealed with a constituent microfibril (MF; 7–10 nm). b) Whole mount of a scalp hair shaft with focal plane on cuticle (Cu, left) and medulla (Md; middle). c) Transverse section of a fine hair shaft revealing a multilayered cuticle (Cu) encompassing a tightly packed cortex (Co) containing many dark melanin granules. From Tobin <i>et al.</i> 2006 [74].	57
Figure 5.3	Hair prepared for microscopic analysis.	60
Figure 5.4	Screenshot of the Olympus DP Manager used to take images	63
Figure 5.5	Screenshot of the Automontage software.....	63
Figure 5.6	– Montaged image.....	64
Figure 5.7	Image of a black medulla that was later removed with Adobe Photoshop.....	64

Figure 5.8 – Extracted image of a ‘Light Brown’ hair.....	64
Figure 5.9 – Extracted image of a ‘Blonde’ hair.....	64
Figure 5.10 The quadratic relationship between microscopic L* measurements and L* measurements by reflective spectrophotometry (RS). The trendline has the equation $y = 0.0188x^2 - 1.2854x + 48.505$ and $R^2 = 0.8164$	66
Figure 5.11 The standard deviation of the intensities of the monochrome hair images compared to the mean intensity value by individual (A) and by image (B).	69
Figure 6.1 The structures of the eumelanin and pheomelanin polymers [76].	72
Figure 6.2 The biosynthetic pathways to eumelanin and pheomelanin as established in human skin and in murine hair pigmentation [76].	73
Figure 6.3 Standard curve for PTCA ($R^2 = 0.9995$) where $y = 34.301x - 138.04$	82
Figure 6.4 Standard curve for PDCA ($R^2 = 0.9951$) where $y = 22.499x - 192.71$	83
Figure 6.5 Sample Chromatograms	84
Figure 6.6 Sample Chromatograms	85
Figure 6.7 Results of the test for linearity of PTCA recovery.	87
Figure 6.8 The amount of PTCA found in 1 mg of hair compared to the RS L* value. The R^2 for the linear relationship of the entire sample set is 0.321 and the R^2 for the circled group is 0.6889. (p-values = 0.000)	88
Figure 6.9 The amount of PTCA found in 1 mg of hair compared to the RS b* value ($R^2 = 0.6561$, p-value = 0.000).....	89
Figure 6.10 The amount of PTCA found in 1mg of hair compared to the Mc L* value ($R^2 = 0.8769$, p-value = 0.000).....	90
Figure 6.11 The mean thickness of the nine hair measurements compared to the amount of PTCA recovered from 1mg of hair ($R^2 = 0.323$, p-value = 0.000).....	91
Figure 6.12 The mean content of PTCA for the LL5 reflective spectrophotometry clusters. Non-Europeans are referred to as Cluster 6 (but were not part of the LL5 analysis). The error bars represent the standard deviation in that cluster.	92
Figure 7.1 Linkage disequilibrium plot of SNP markers that have been mapped according to their chromosomal locations.	105
Figure 7.2 The dominant genotype model for association with a* for rs1805007 (MC1R; R151C). D represents the minor allele and d represents the major allele.	114
Figure 7.3 The recessive genotype model for association with a* for rs1805007 (MC1R; R151C). D represents the minor allele and d represents the major allele.	115
Figure 7.4 The associations between the rs16891982 and rs1805005 dihybrid and the LL5 Blonde/Fair Cluster 2.....	118
Figure 7.5 The associations between the rs16891982 and rs1805005 dihybrid and L* value ..	119
Figure 7.6 The associations between the rs16891982 and rs1805005 dihybrid and b* value ..	119

LIST OF TABLES

Table 3.1 The self-reported hair colours (SRC) compared to the observer-reported hair colours (ORC) of individuals with European ancestry.....	30
Table 3.2 Composition and cluster characteristics of five clusters determined by the log-likelihood probability method (LL5).....	33
Table 3.3 Composition and cluster characteristics of two clusters determined by the log-likelihood probability method (LL2).....	33
Table 3.4 Composition and cluster characteristics of six clusters determined by the log-likelihood probability method (LL6).....	33
Table 3.5 Observer reported colours found in the LL5 clusters and descriptions of individuals compared to other ORC group members.	35
Table 4.1 Percent of cases correctly classified by a discriminant analysis on Clusters determined by RS or the observer reported colour group (ORC), using DI data.....	51
Table 6.1 Variability in the hair degradation and HPLC method	86
Table 7.1 Genes involved in the synthesis of melanin.	96
Table 7.2 Genes involved in melanosome structure and transport.....	96
Table 7.3 Genes involved in the type of melanin produced (Eumelanin/Pheomelanin switch). ..	97
Table 7.4 Transcription factors that may be involved in human pigmentation.....	97
Table 7.5 Single Nucleotide Polymorphisms tested for association with hair colour and the genes in which they are found.....	100
Table 7.6 SNP results showing the minor allele frequency and the call rate – non-polymorphic SNPs have a minor allele frequency of 0.00.....	103
Table 7.7 SNPs found to be significantly associated with LL5 Red (Cluster 1) at the $\alpha = 0.1$ level for the allelic model (N = 8)	106
Table 7.8 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the allelic model (N = 32)	107
Table 7.9 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the dominant model (N = 32).....	107
Table 7.10 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the recessive model (N = 32)	107
Table 7.11 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the allelic model with Cluster 1 removed (N = 32)	108
Table 7.12 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the dominant model with Cluster 1 removed (N = 32).....	108
Table 7.13 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the recessive model with Cluster 1 removed (N = 32)	108
Table 7.14 SNPs found to be significantly associated with LL5 Dark (Cluster 5) at the $\alpha = 0.1$ level for the allelic model (N = 86)	109
Table 7.15 SNPs found to be significantly associated with LL5 Dark (Cluster 5) at the $\alpha = 0.1$ level for the dominant model (N = 86).....	109
Table 7.16 SNPs that have significant allelic associations with L* (D vs. d).....	110
Table 7.17 A test for genotypic associations with L* (DD vs. Dd vs. dd)	111
Table 7.18 SNPs that have a significant association with L* in dominant model (DD, Dd vs dd)	111
Table 7.19 SNPs that have a significant association with L* in a recessive model (DD vs Dd, dd)	112
Table 7.20 SNPs that have significant allelic associations with a* (D vs. d).....	112
Table 7.21 A test for genotypic associations with a* (DD vs. Dd vs. dd)	113
Table 7.22 SNPs that have a significant association with a* in dominant model (DD, Dd vs dd)	113
Table 7.23 SNPs that have a significant association with a* in a recessive model (DD vs Dd, dd)	113
Table 7.24 SNPs that have significant allelic associations with b* (D vs. d).....	116
Table 7.25 A test for genotypic associations with b* (DD vs. Dd vs. dd)	116
Table 7.26 SNPs that have a significant association with b* in dominant model (DD, Dd vs dd)	117
Table 7.27 SNPs that have a significant association with b* in a recessive model (DD vs Dd, dd)	117

Chapter 1 Introduction

1.1 General Introduction

1.1.1 Introduction

“Color is so much a matter of direct and immediate perception that any discussion of theory needs to be accompanied by experiments with the colors themselves.”

-- Walter Sargent (1868-1927)[1]

Pigmentation in humans has been observed and studied for at least 4000 years [2] and is an integral part of our physical health and cultural identity. Scientists have been interested in the inheritance of pigmentation traits for over a century and have been studying the genetics of pigmentation for more than 50 years [3]. Efforts in the past have concentrated a great deal on pigmentary disorders [4, 5] and on murine models of pigmentation, where over 150 alleles at more than 90 genetic loci have been discovered so far [6].

Study of the normal pigmentation variation in humans is more recent and has been investigated by medical scientists [7], anthropologists [8] and, more recently, by forensic science researchers for the prediction of visible phenotypes to be used as an investigative tool [9]. In the field of forensic science, iris colour has received a great deal of attention, as has the genetic determination of biogeographical ancestry (BGA) [10-12]. An American company, DNAPrint, has developed tests for BGA that have been used in casework in the United States [12] and are developing a kit for the prediction of iris colour [9, 13].

Hair colour has received much less attention in the study of pigmentation, with most work on hair colour being done as a consequence of, or in correlation with, another pigmentary trait (such as iris colour or melanoma risk phototype). Parts of this investigation have been done by researchers in the vastly different fields of medical science [7, 14, 15], the cosmetics industry [16, 17] and forensic science [18], as well as in chemistry and physics [19, 20]. The reported project is specifically concerned with genetic determinants of blonde hair colour.

Many parties may be interested in research on blonde hair colour. Blonde hair colour has been implicated as a melanoma risk factor [7] and genetic determinants may be interesting to medical, anthropological and cosmetic researchers. The prediction of physical traits for use in forensic investigations is a primary interest in this research and hair colour is one of many traits that may assist investigators in limiting a suspect pool. The UK Forensic Science Service has already implemented a test with 84% accuracy for predicting red hair [18] .

To produce an accurate and valid test for physical characteristics, the genetic basis of these characteristics must first be solidly understood. The vast variety of normal pigmentation variation is a huge genetic puzzle and hair colour is one part of this puzzle. By sequentially examining smaller components, the larger picture of human pigmentation variation will emerge.

1.1.2 Research Outline and Objectives

It is hypothesised that there are genetic variants specifically associated with an objectively defined Blonde phenotype. This project aims to first characterise and define the phenotype under study ('blonde hair') and then to evaluate DNA sequence variants that may be associated with this phenotype. These variants may be useful in predicting the hair colour of an unknown subject.

A review of the scientific literature relevant to forensic science and the prediction of physical features can be found in Chapter 2. Further discussion of the background literature that is relevant to each experiment can be found in Chapters 3 through 7.

As well as incorporating a review of the applicable literature, Chapters 3 through 7 discuss the methods and results of each experiment. Chapters 3 and 4 are concerned with objective measurement and definition of hair colour phenotypes. Chapters 5 and 6 involve a more thorough characterisation of a subset of samples and Chapter 7 describes the genetic analysis.

In Chapter 3, natural hair colour is measured by reflective spectrophotometry and these objective measurements are compared to observer and self-reported colours and are

classified into defined groups by cluster analysis. Chapter 4 involves the comparison of the reflective spectrophotometry measurements with colour measurements obtained by digital image analysis.

Objective 1 - Examine the natural, adult hair colour variation among a population of European ancestry and evaluate methods for both the measurement and classification of hair colours.

Why? To assist in the development of standardisation in hair colour research for this project, for future research and in forensic science.

Chapter 5 reports the findings of a microscopic hair examination of a subset of individuals and compares microscopic features of hair colour and hair texture with the macroscopic colour. A chemical analysis of the pigments in hair is described in Chapter 6 and the chemical composition of hair pigments is compared to macroscopic and microscopic colour measurements.

Objective 2 – Characterise the microscopic and chemical features of hair colour in this sample and how they relate to the macroscopic colour.

Why? To increase the understanding of hair colour biology and chemistry; this will assist in both traditional forensic hair examination and in designing and interpreting genetic studies.

Chapter 7 examines the relationship between hair colour and genotype at the site of 58 single nucleotide polymorphisms. Associations within the sample are presented and additional findings will be discussed.

Objective 3 – To compare genetic results with objectively measured and defined hair colour phenotypes in the search for genetic associations with ‘blonde’ hair colour.

Why? Having this information will assist in developing forensic predictive tests for physical features.

A final discussion of the research, its findings and possible future directions will be presented in Chapter 8.

Chapter 2 Background and Literature Review

2.1 Forensic Biology

2.1.1 Forensic Hair Examination

Hair is commonly encountered as physical evidence at crime scenes [21, 22]. Traditional forensic hair examination involves examining the specimen under a microscope and observing characteristics such as colour, length and diameter of the hair, as well as the distribution of pigment granules and other morphological features of the hair that may assist in determining the source of the hair [21-24].

An expert hair examiner may be able to determine a great deal of information about the contributor of a hair and will compare the features of a ‘questioned’ hair to those of a hair sample from a ‘known’ source. The examiner may be able to determine if the hair is human and if not, what species it belongs too [22]. Human hair will be evaluated for its length and shape, which will help determine the racial origin of the hair and the area of the body it is from. The tip of the hair (e.g. cut or abraded), the macroscopic colour of the hair and any evidence of disease will also be evaluated [22].

Microscopic features of the hair will also be examined. This will include evaluations of the hairs’ colour, diameter, cross-section, pigment features, medulla classification, cortical features and cuticle features [22]. An in-depth description of forensic hair examination can be found in James Robertson’s book *‘Forensic Examination of Human Hair’* (1999) [22].

Figure 2.1 gives an example of checklists that is used by the Australian Federal Police forensic hair examiners.

AUSTRALIAN FEDERAL POLICE

FORENSIC SERVICES, SCIENTIFIC

Page ____ of ____

EXAMINATION RECORD

Case Reference: _____

Hair Examination Record - Sheet 1

HAIR NUMBER	1	2	3	4	5
MACROSCOPIC FEATURE	r t	r t	r t	r t	r t
LENGTH cm					
SHAFT PROFILE Straight Wavy Curly Peppercorn					
COLOUR ¹ Colourless Yellow Brown Reddish Black					
ROOT Absent Club Ribbon Sheath Present					
TIP Natural Taper Cut Rounded Frayed or Abraided Split Crushed or Broken Singed					

General Description and Comments:

- ¹ A assessed with a stereo microscope using standardised illumination
- B basic colour to be qualified by shade or depth of colour. Light (L), Mid (M) or Dark (D)
- C note artificial colouring

Examined by: _____

Notes by:

Day:

Date:

Time:

AUSTRALIAN FEDERAL POLICE						
FORENSIC SERVICES, SCIENTIFIC			Page ____ of ____			
EXAMINATION RECORD – Sheet 2		Case Reference: _____				
HAIR NUMBER		1	2	3	4	5
MICROSCOPIC FEATURE		r t	r t	r t	r t	r t
SHAFT DIAMETER	max units at × 40					
PIGMENT DENSITY	None Light Medium Heavy Opaque					
PIGMENT DISTRIBUTION (across hair shaft)	Uniform Towards Medulla Towards Cuticle To One Side					
PIGMENT AGGREGATE SHAPE	Streaked Clumped Oval Clumped Round					
PIGMENT GRANULE SHAPE	Fine Oval/Oblong					
PIGMENT AGGREGATE SIZE	Fine Medium Coarse					
PIGMENT GRANULE SIZE	Fine Medium Coarse					
OVOID BODIES ²						
MEDULLA DISTRIBUTION	None Medulla < Space Medulla > Space Continuous					
MEDULLA TYPE	Opaque Translucent					
CORTICAL FUSI ²						
CORTICAL TEXTURE	Not Visible or Smooth Visible or Coarse					
CUTICLE	Thickness units × 40 Colour					
CUTICLE OUTER	Smooth Serrated Ragged Cracked Looped					
² Where present their shape, size and distribution both along and across the hair shaft may have some value when used as comparative features.						
Examined by: _____			Checked by: _____			
Notes by:	Day:	Date:	Time:			

Figure 2.1 Hair examination checklists for the Australian Federal Police

The American National Research Council recently published a document entitled “Strengthening Forensic Science in the United States: A Path Forward” [25]. This document expresses the need for standardised terminology and reporting of results in forensic sciences, as well as for an improved scientific basis for forensic science techniques and knowledge [25]. In contrast to DNA evidence, which uses standardised and objective protocols to determine genotype data and then uses a statistical calculation for identification, traditional hair examination still involves an expert’s opinion and a hair cannot currently be conclusively attributed to an individual. The features that are highly variable between individuals are those which are difficult to capture easily or objectively, such as pigment pattern and colour. However; there are efforts being made to develop objective and validated methods for the forensic examination of hair and to increase the value of hair as evidence [24, 26]. One of these efforts involves the use of pigmentation patterns and colour measurements for identifying the contributor of a hair [27].

2.1.2 DNA

Deoxyribonucleic acid, or DNA, is a molecular instruction book that resides in almost every cell of our body. This molecule contains the information about what blood type a person is, their predisposition for many diseases and, theoretically, what one looks like. The DNA in each individual is unique, with the exception of identical twins. This uniqueness allows for a powerful means of determining the source of biological material found at crime scenes or the identification of victims of crime. DNA identity testing has become routine in the course of criminal investigations [28, 29]. By comparing to a sample of known origin, forensic biologists determine the probability that an individual was the source of a particular DNA sample; whether it is a blood stain, skin under the fingernails of a victim or the paternity of a child. The word “match” is used in conversation, however; it is not quite correct. A more correct phrase would be to conclude that a subject ‘can not be excluded as the contributor of the DNA’ and the likelihood of someone else being the contributor is expressed as a probability [28]. A full, good-quality, DNA profile produces probabilities such as one in trillions – a conclusive result.

This profile is typically arrived at by the testing of genetic markers called Short Tandem Repeats (STRs). Short Tandem Repeats involve regions of the genome that do not code for specific proteins (non-coding regions) and where short sequences of nucleotides are repeated. For example, the STR called TH01 is used in forensic science and consists of five to eleven repeats of the sequence AATG [21, 28]. As for any genetic marker, there are two copies; one on the maternal chromosome and one on the paternal chromosome, which may have the same or a different number of repeats. Therefore, in this example, an individual may have, for that marker, a TH01 profile of 6 and 8 repeats.

STRs are typically tested by capillary electrophoresis in the laboratory via strictly controlled and validated protocols where the size of the markers (and therefore, the number of repeats) is determined [28]. The number of markers tested will depend on the case involved and the exact markers tested will depend on the laboratory. A well known system is the Combined DNA Index System (CODIS) used by the FBI. This consists of 13 STR markers that are completely independent of one another and that have known population frequencies [28]. In addition to the STR marker, one coding

sequence is typically used in forensic DNA profiling to determine the sex of the individual. The amelogenin gene (which codes for an extracellular matrix protein in tooth enamel) is shorter by six base pairs on the X chromosome compared to the Y and so a female will show one fragment size, a male will show two [28]. Australian forensic scientists use a similar system for the National Criminal Investigation DNA Databank (NCIDD at www.crimtrac.gov.au).

It is important to note that in normal DNA profiling for forensic purposes, only non-coding genetic markers are used (except for amelogenin) and, besides sex, nothing is known about the individual, it is only a test for identity. In such profiling, samples from suspects, criminal databases and victims are compared to samples from the scene of a crime. If there is no comparison sample available (such as a suspect sample or database hit), the DNA sample is of limited investigative value.

A typical DNA profile of a female using nine STRs and amelogenin is shown in Figure 2.2.

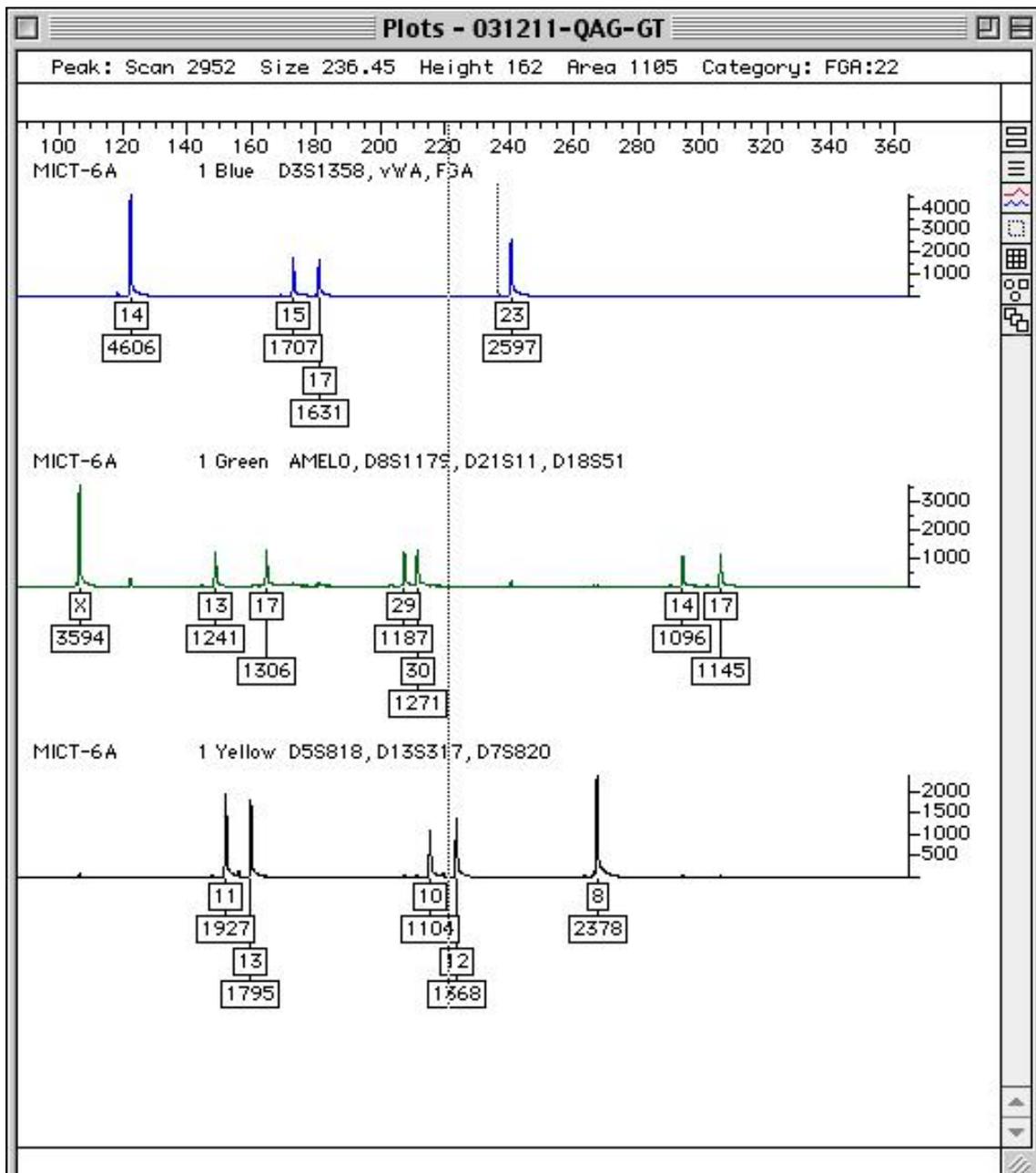


Figure 2.2 A nine-locus (+ amelogenin) profile of a female that may be used for identification

2.1.3 Future Directions in Forensic Science

The field of forensic science is changing rapidly, with heightened efforts in scientific research as a response to increased demands from the courts and investigators for useful, validated techniques. Evidentiary forensic science involves the comparison of samples of ‘questioned’ origin to those of ‘known’ origin; whether it is a DNA sample, a hair or other trace evidence or a tool mark impression.

A newer area of forensic science is ‘forensic intelligence’ – forensic science that assists with an investigation; prior to an arrest or court-room activity [30]. Science may be able to provide the police with information that can direct them to a particular individual or location. Some examples of this in use today include psychological and geographic profiling as well as the linking of crime scenes through trends in physical evidence such as foot prints, tool marks and DNA [30].

Using genetic information to predict what someone looks like is one area of investigative forensic science that has gained a great deal of interest in recent years. This is alternatively called ‘phenotype profiling’, ‘physical feature prediction’ or ‘molecular photofitting’. Phenotype profiling is based on the idea that our physical features (such as skin, hair and eye colour, baldness patterns, height, etc.) are coded for in our DNA and if one was to test for these variations, a description of an individual could be produced and offered to investigators. Having this information would assist the police in narrowing a suspect pool or beginning a search for a perpetrator, a victim or even a witness [9, 31].

2.2 Phenotype Profiling

2.2.1 Ancestry

One of the more developed areas of research in phenotype profiling is the estimation of a person’s biogeographical ancestry (BGA) from their genetic sequence [9]. Ancestry Informative Markers (AIMs) are genetic markers (usually single nucleotide polymorphisms or SNPs) that vary in frequency between populations that were once geographically isolated from one another due to genetic drift [9]. By genotyping many of these markers a statistical estimate of a person’s genomic ancestry can be determined. The strength of the estimate may depend on how many markers are used in the calculation and how informative the particular markers are for ancestry [9, 32, 33].

It is important to differentiate between genealogical ancestry and genomic ancestry. Genealogical ancestry depends on the population one’s ancestors derived from (fixed by history) where genomic ancestry is the average level of *admixture* of an individual [9]. For example; a person may have 3 grandparents of European origin and 1 of African

origin. Therefore their genealogical ancestry would be 75% European and 25% African, but their genomic ancestry (or BGA) may not be the same. Due to independent assortment and recombination of chromosomes, this person may have more or less than or equal to 25% African genomic ancestry. Figure 2.3 illustrates a hypothetical example where an individual has 75% European and 25% African genealogical ancestry, but approximately 61% European and 39% African genomic ancestry due to random assortment of chromosomes in meiosis. This example does not take into account any recombination of chromosomes.

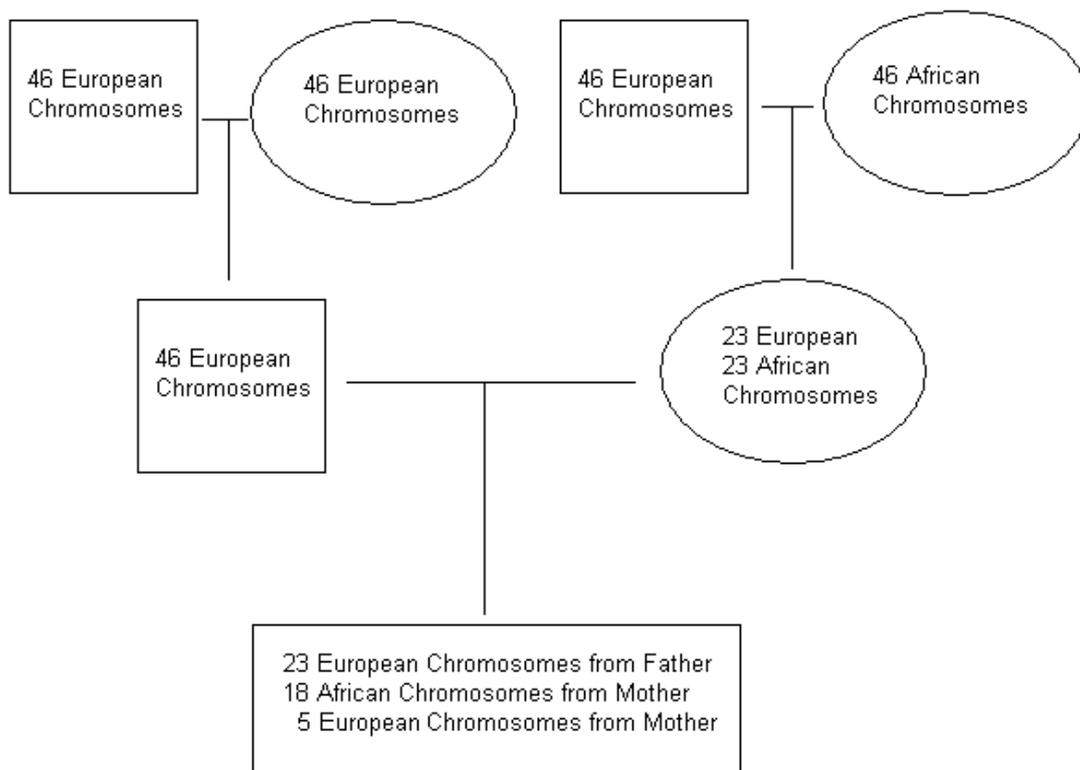


Figure 2.3 A hypothetical pedigree of one individual who has 75% European and 25% African genealogical ancestry, but approximately 61% European and 39% African genomic ancestry due to random assortment of chromosomes in meiosis. This example does not take into account any recombination of chromosomes.

Accurate measures of biogeographical ancestry are necessary to assist clinicians, biomedical researchers and forensic scientists [9]. There is often variation in drug responses across various populations and knowledge of a patient's BGA will assist a clinician in determining the correct treatment [9]. Assessment of BGA also increases the accuracy of medical genetic studies by sorting out which variants are related to ancestry rather than disease [9] (this may also apply to the study of complex non-

disease traits such as pigmentation). Forensic DNA analysts are interested in measures of BGA to assist with likelihood calculations for identity testing and for inferring physical characteristics of an individual [9].

2.2.2 Indirect Phenotype Inference

Physical traits such as pigmentation and craniofacial morphology can be inferred from ancestry using an indirect method of phenotype inference [9]. In this case, ancestry, as calculated by testing AIMS, is used as a proxy for the net effect of phenotypically functional loci and can give information on traits that have a strong association with ancestry, such as pigmentation [9]. For example, an individual with 80% African ancestry is likely to have darker skin than an individual with 30% African ancestry, however, the range of informative admixture values must be determined empirically [8, 9]. Frudakis (2008) discusses the use of admixture databases to assist with the interpretation of indirect phenotype inference [9].

2.2.3 Direct Phenotype Inference

Direct phenotype inference involves the evaluation of genetic information that is directly related to the trait being examined [9]. This can be used if the genetic cause of the trait is understood and results in higher quality information than an indirect inference, however this information is much harder to come by [9].

Pigmentation may have a *relatively* more straight forward genetic mechanism than something like cardiovascular health due to more obvious selection pressures and this makes it more amenable to study the direct genetic causes of pigmentation traits [9], however there is still a great deal of information to be gathered and understood.

To develop a predictive test for a pigmentation feature, Frudakis (2008) presents four steps that may be considered. First, associations must be identified between genetic markers and phenotypic traits. Next, associations between certain genetic polymorphisms and certain trait values must be established. The inferential method for predicting an individual's trait value must be then validated and finally, the trait can be inferred directly [9].

The presented research is primarily concerned with the first step in developing method for the direct inference of hair colour from genetic information (finding associations) and the literature in this field will be discussed next.

2.3 Direct Inference of Hair Colour

2.3.1 Human Hair Colour

Hair colour is one physical feature of interest to those studying the genetics of physical traits for phenotype profiling. In humans, hair colour is a continuous trait [9], but one that is usually broken into categories such as “blonde” or “brown” for reference. Hair colour in humans is the result of two main pigments, the black eumelanin and red pheomelanin, being transferred from the melanocytes where it is produced to the keratinocytes of the hair shaft. Details of the pigmentation process (as it is thought to occur in the epidermis) and the genes thought to be involved will be discussed in later sections of this thesis; however an overview of the process is illustrated in Figure 2.4.

The final colour of the hair depends on the amount and type of melanin, which are related to the shape, size and type of melanosomes (the organelles where melanogenesis occurs) and the activity of Tyrosinase (the primary enzyme responsible for melanin production) [9]. One can see from Figure 2.4 that there are many genes involved in the control of this process and there are likely more genes that have not yet been associated with pigmentation at all.

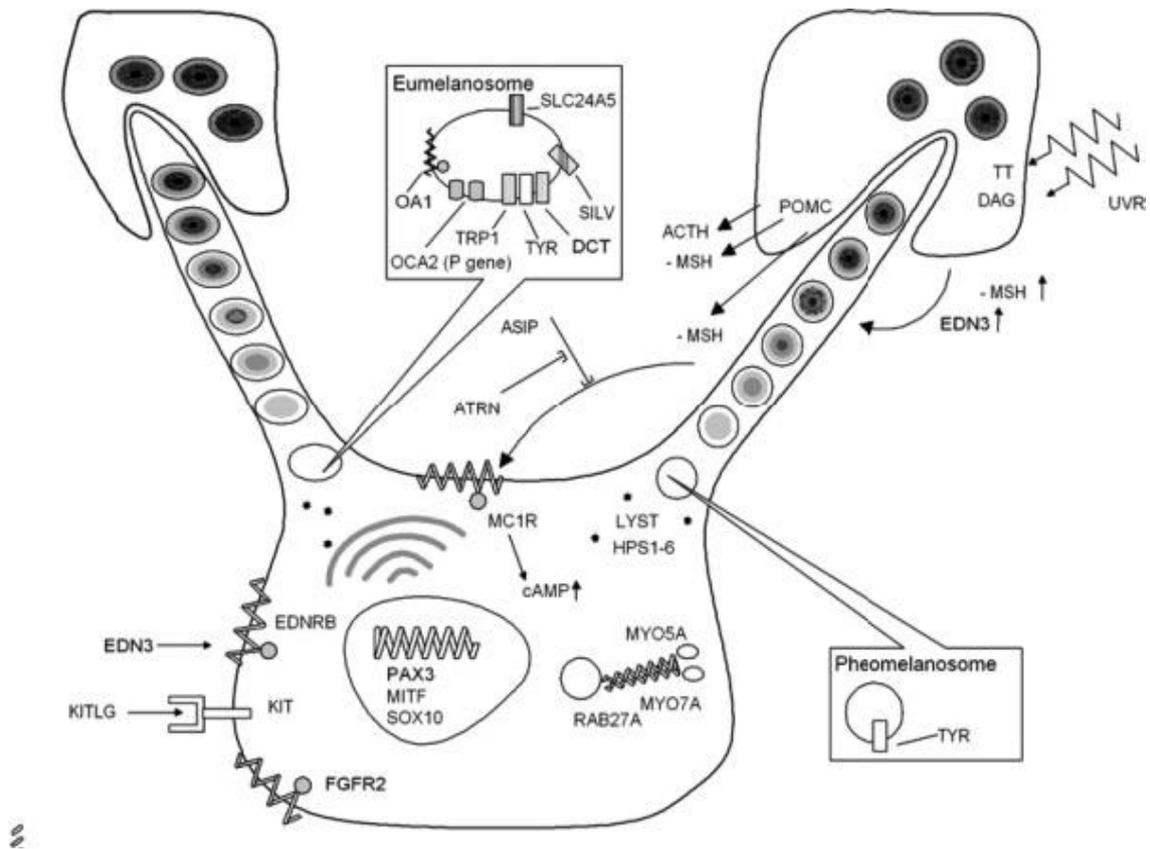


Figure 2.4 Graphical representation of a melanocyte showing some important proteins involved in melanocyte development and the pigmentation pathway (as thought to occur in the epidermis), including the following: 1) the tyrosinase enzyme complex (TYR, TRP1, DCT) located on the membrane of the melanosome, which is responsible for the enzymatic conversion of the amino acid tyrosine into melanin; 2) other proteins located within the melanosomes that play a critical role in melanogenesis (MATP, P gene, SILV, SLC24A5); 3) the signalling pathways affecting the regulation of melanin synthesis, including hormones and receptors (α -MSH, MC1R, ASIP, ATRN); 4) the transcription factors involved in melanin production (PAX3, MITF, SOX10); 5) proteins involved in melanosome transport (MYO5A, MYO7A, RAB27A) and melanosome construction/protein routing (CHS1, HPS1-6); and 6) developmentally important ligands (EDN3, KITLG) and receptors (KIT, EDNRB) which control melanoblast migration and differentiation [34].

Research into finding genetic associations with hair colour has been less successful and is a less developed field than research in the areas of skin and iris pigmentation. Iris pigmentation is thought to be primarily associated with variations the OCA2 gene (also known as the P gene) shown in Figure 2.4 [9, 10, 35]. Iris colour has been well characterised, phenotype data bases have been created and luckily, while several genes are likely involved in iris pigmentation the OCA2 gene seems to have the greatest effect [9].

There are several possible reasons that the search for genetic associations with hair colour has been less successful and Section 2.3.3 will discuss these issues. They are also summarised in Frudakis (2008) [9].

2.3.2 Genetic Studies of Hair Colour

The first association of a genetic variation with hair colour was in 1995 where variations within the gene MC1R, which codes for the melanocortin 1 receptor (the melanocyte-stimulating hormone receptor or MSHR), were found to be associated with red hair [36]. The interest in red hair and MC1R continues [14, 37-42], perhaps due to its associations with skin-cancer risk and perhaps due to the fact that it is reasonably simple to study (MC1R is a small gene with very clear effects). Due to an obvious phenotype and a simple genetic mechanism, testing for red hair has also become one of the first direct phenotype inferences used in forensic science [18].

Research into the genetics of non-red hair colours and genes other than MC1R is less well established. One polymorphism in the 3' untranslated region of the human agouti signalling gene (ASIP g.8818A>G) has been associated with darker pigmentation in African populations [43] and with darker hair and eye colour in Europeans [44]. Two polymorphisms in the MATP gene – also known as SLC45A2 - (Phe374Leu and Glu272Lys) have been associated with darker pigmentation (including hair) in European populations [45].

Recently (December 2007), an Icelandic company called 'deCODE genetics' published a very large study on the genetics of various pigmentation traits, including hair colour [46] which found several associations with Red hair, as well as Blonde (vs. not-Blonde) in genes such as MC1R, OCA2, TYR, KITLG and SLC24A4.

2.3.3 Issues in the Study of Human Hair Colour Genetics

One major issue in the study of human hair pigmentation is that different genes may be involved than are involved in iris and skin colour [9]. Pigmentation has traditionally been studied in mice [47, 48] and in human epidermis [4]. It is now known that there are fundamental differences between mouse and human pigmentation and between skin and hair pigmentation. For example, pheomelanin, the red-yellow pigment that is found in high proportions of red-haired humans, sometimes appears yellow in mice and these mice (*recessive yellow*) have almost no eumelanin [47], which does not occur in humans (see also Box 6.1) [20, 47].

The differences between skin and hair pigmentation are also being elucidated. It is well known that a person can have dark hair and fair skin and sometimes even fair hair and dark skin (such as is found amongst some Melanesian and Australian aboriginal populations) [49]. This is possibly the result of the process of melanin degradation being different in hair and skin and may be due to other means of control [49]. Recent gene expression studies [50] have shown that at least one protein (Tyrosinase-related protein Type 2) is expressed in epidermal melanocytes, but not in hair follicle melanocytes.

In addition to different genetic mechanisms, hair colour may be affected by the environment (external and internal) more so than skin or iris pigmentation [9]. It is widely known that hair colour can change throughout a person's life. Hair colour often changes in puberty and as a person ages. Hormonal regulation of melanogenesis is thought to be involved in these processes [51] and hormonal disorders or certain medications may have additional effects on hair colour. Hair may also change in response to external environmental conditions, such as sunlight. Photo-lightening of hair has a great deal of anecdotal evidence, but few scientific studies. Takahashi and Nakamura (2004 and 2005) measured the photo-lightening of blonde and red hairs, and found that strands of both lightened in response to visible and ultraviolet light, by a maximum of about five ΔL^* units [16, 17] (ΔL^* and the CIE $L^*a^*b^*$ space will be discussed in Chapter 3). Larger studies on heads of hair rather than strands may assist in quantifying the photo-lightening affect of sunlight on hair.

Hair may also be coloured artificially and a person's memory may not be accurate when asked to self-report hair colour later on. Each of these factors may introduce error into genetic studies of hair colour and will present problems if predictive tests are put into use in the future; however these factors can be controlled for in studies at this point in the research process by choosing subjects who are healthy and have their natural, adult hair colour at the time of sampling.

Unknown genetic interactions may also complicate the study of hair colour. There is most likely epistatic interactions between genes involved in hair pigmentation and they may be different than any discovered in epidermal or iris pigmentation. For example, a mutation in the MATP gene (g.8818A>G) is thought to suppress red-hair causing mutations in MC1R and leave an individual with darker hair [44].

Another major obstacle in the genetic study of hair colour, as previously mentioned, is the difficulty in defining the hair colour phenotypes. Studies of hair colour often ask the subject to report their own hair colour or the observations of the researcher are used [38, 45, 46] and while objective measurements are becoming more common, they are sometimes weighted towards one hair colour [52] or involve children, whose hair colour may yet change [41]. Objective hair colour measurements are difficult to obtain on large samples; however standardisation in methods of hair colour description and classification will increase comparability of various research studies and will assist in communication between research groups, forensic scientists and investigators.

The relationship between genetic variations and hair colour may also be more closely related with different types of hair colour measurements [53], such as chemical composition or microscopic colour and understanding the relationships between these measurements and macroscopic hair colour will assist in defining the hair colour phenotypes.

As mentioned above, the deCODE genetics study found several associations with Red hair, as well as Blonde (vs. not-Blonde) in genes such as MC1R, OCA2, TYR, KITLG and SLC24A4. This study lends support to the idea that several, less penetrant alleles are responsible for non-red variation. The odds ratios (a measure of the effect a mutation has) for blonde hair are lower than what is typically found for associations with red hair. The authors (Sulem *et al.*) note that no objective measurements of hair

colour were carried out. This study included almost 7000 individuals and illustrates two problems in analysing hair colour genetics – that hair colour is likely controlled by several low-penetrance alleles (which are harder to find in a case-control study) and that objective measurements of hair colour (for consistency and accuracy) are difficult or impossible in such large studies. In the future, it may be possible to extrapolate information from the measurement of smaller data sets, such as that presented in this project, by constructing a visual database of hair colour groups.

2.3.4 Continuing the Study of Hair Colour

As outlined in Section 1.1.2, each of the following chapters measured hair colour at a different level or in a different way. Further literature, relevant to a specific experiment, will be included in each chapter. Macroscopic colour, microscopic colour, chemical composition and genetic variation will be examined, analysed and discussed and it is hoped that each section individually and when considered together will provide information on hair colour that will be interesting and useful to those in the fields of genetics and forensic science.

Chapter 3 Hair Colour Measurement and Variation¹

3.1 Introduction

3.1.1 Colour and Perception

Colour is defined by the Random House Unabridged Dictionary as “the quality of an object or substance with respect to light reflected by the object, usually determined visually by measurement of hue, saturation, and brightness of the reflected light”[54].

When light of a certain wavelength is reflected from an object (the dominant wavelength), it appears to the human observer as colour [55]. Visible light or “white light” is made up of electromagnetic radiation wavelengths in the 380 nanometre to 780 nanometre range. Figure 3.1 illustrates how these colours relate to other forms of electromagnetic radiation and what colours result from the visible spectrum.

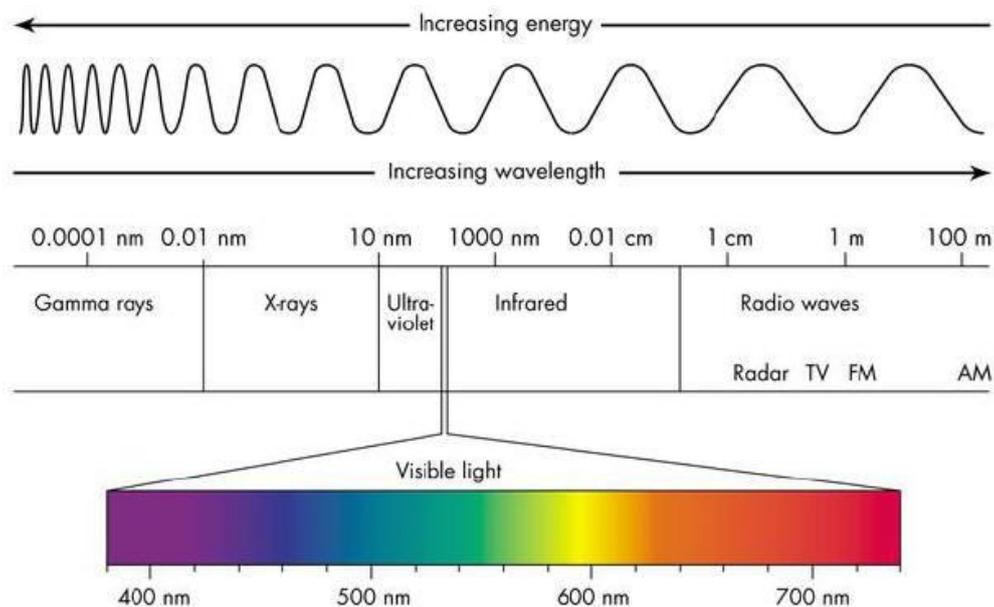


Figure 3.1 Energy spectrum including the wavelengths of visible light that appear as colours²

The human eye is an incredibly complex structure and is capable of detecting wavelengths of light in the visible spectrum [55]. For the purposes of this research, the

¹ An earlier version of this experiment was published in the American Journal of Physical Anthropology Vaughn, M. *et al.* 2008: 137(1) Pages 91-96. Please see Appendix 4

² http://www.antonine-education.co.uk/physics_gcse/Unit_1/Topic_5/em_spectrum.jpg

colour sensitivity of the eye will be briefly discussed. In the retina (which covers most of the interior of the eye) there are photosensitive receptors called rods and cones that feed stimuli to the relevant neurons and nerves. Rods give monochromatic vision while cones contribute to colour vision [55], although all contribute to three ‘channels’ of vision; *achromatic* (brightness) and two *chromatic* channels (red-green and yellow-blue) [55, 56]. This is referred to as having a *trichromatic* system of colour vision.

3.1.2 Colour Measurement

To objectively measure hair colour, a method of measuring colour must first be chosen. As mentioned in the above definition, colour can be described in terms of its *hue* (H), *saturation* (S) and *brightness value* (V) [55] as shown in Figure 3.2. This type of model is referred to as a tri-stimulus colour model or colour space. Other examples include the additive *Red-Green-Blue* (RGB) colour model that is often used to display images on a computer screen (Figure 3.3) and the subtractive *Cyan-Magenta-Yellow* (CYM) model (Figure 3.4) that is found in some printers [57]. RGB and CMY colour models are useful, however are difficult to associate with human perception of colour and are mainly used for devices.

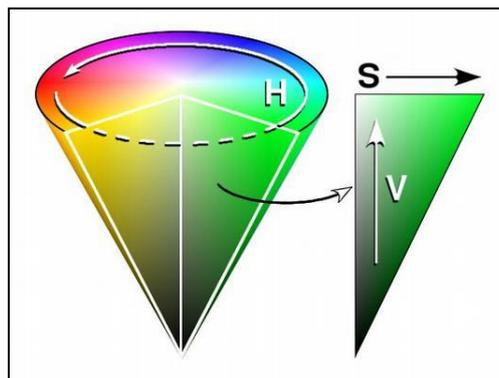


Figure 3.2 The HSV colour space showing hue (H), saturation (S) and brightness value (V)

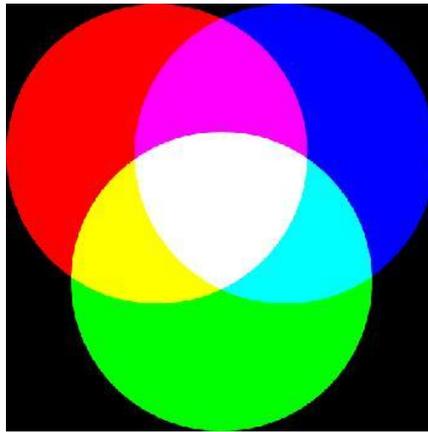


Figure 3.3 Diagram of the Red-Green-Blue (RGB) colour model.

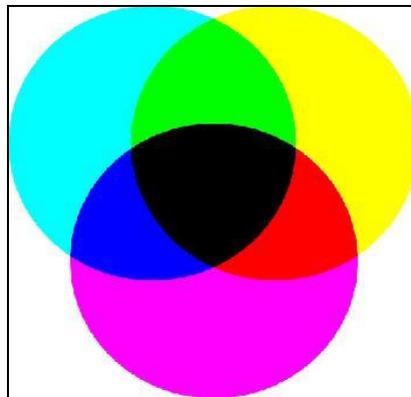


Figure 3.4 Diagram of the Cyan-Magenta-Yellow (CMY) colour model.

Another tri-stimulus colour model is the CIE $L^*a^*b^*$ model (Figure 3.5). The International Society of Illumination (Commission Internationale de l'Eclairage) has developed several similar colour models over the years, based on human perception, and the $L^*a^*b^*$ model is the one that seems to best represent how we think about hair colour. In this model, colour is measured on three axes that correspond to the trichromatic model of human perception [58].

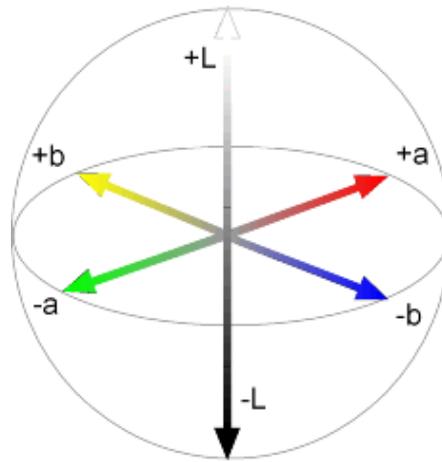


Figure 3.5 An illustration of the CIE L*a*b* colour space showing the three axes: L* (Black = 0 to White = 100), a* (-100 = green to +100 = red) and b* (-100 = blue to +100 = yellow).[57]

The lightness, or intensity, of a colour is measured on the ‘L*’ axis on a scale from 0 (being black) to 100 (being white). The colour is also measured on the ‘a*’ axis that gives a value from -100 (being green) to +100 (being red). The ‘b*’ axis measures colour from -100 (being blue) and +100 (being yellow). This model has the same benefit of the HSV model in that it corresponds to human perception of colour instead of a device-based model and has the additional benefit of giving a grid point for each specific colour [57] and this grid point allows for the mathematical comparison of colour. In addition, this model has been used for several other studies of human pigmentation [16, 17, 34, 59]. It should be noted that, theoretically, the a and b axes have no maximum or minimum values but this research has used the cut off points of +/- 100 because these represent the practical limit of the instrument and software used for colour measurement [60].

The comparison of colours in the CIE L*a*b* system often uses the delta (Δ) symbol for the ‘change’ or ‘difference’ in colour. ΔL^* would be the difference between two L* values while Delta E* (ΔE^*) is an Euclidian measure of the total distance between two CIE L*a*b* colour values and is calculated according to the following formula [55]:

$$\Delta E^* = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}$$

Figure 3.6 illustrates a difference of 5 units along the L*, a* and b* axes and one can see that the differences are discernable, but not obvious. It should be noted that the

colours in Figure 3.6 are approximate and only for illustration purposes, due to the differences in the way various computer screens and printers reproduce colour.

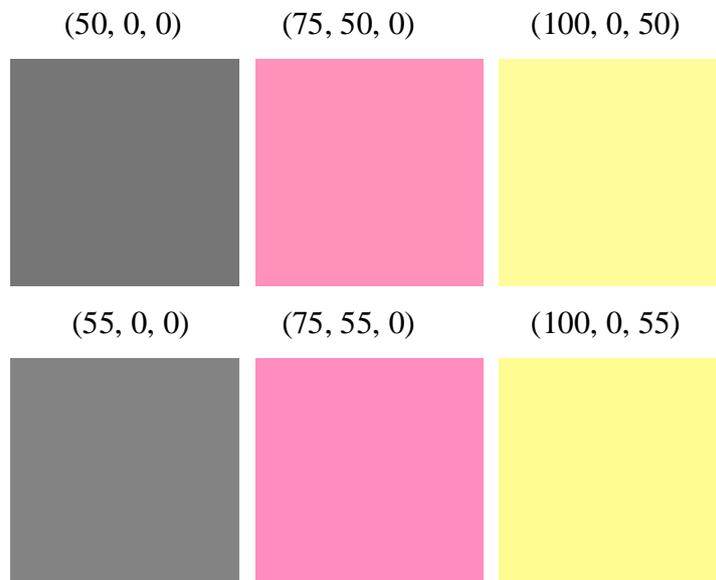


Figure 3.6 An example of a 5 unit difference along the L*, a* and b* axis.

Now that a system of measurement has been defined, a method must now be determined. There are many ways to determine the colour of an object. Reflective spectrophotometry (RS) measures the wavelengths and amount of light reflected from an object and provides its own light source. This gives an accurate and consistent colour reading. Applications of RS include the measurement of pigments for paint and materials manufacturing and for determining food quality [61]. Colour may also be judged by comparing it to a colour chart (e.g. Munsell Book of Color) [55] which, while simple, involves a certain amount of subjectivity on the part of the observer.

Reflective spectrophotometry has been used to objectively measure skin pigmentation [8, 59] and the technique has also been used on hair in different ways and for various purposes including cosmetic science [16, 17], for the determination of hair composition and structure [19], for comparison of measurement methods [59] and in recent genetic studies [41, 52]. It has been found to be an objective and effective way to measure hair colour.

In 2000, Shriver and Parra used the CIE L*a*b* system with reflective spectrophotometry to measure the hair colour of 41 European-American individuals and 18 individuals of non-European ancestry, comparing the L* colour component to the Melanin Index [59]. They found a significant correlation between these systems of measurement and reported low variability in the hair colour of individuals of non-European ancestry. Naysmith *et al.* in 2004 used the CIE L*a*b* system to measure the hair colour of 50 individuals for genetic and chemical studies of red hair [52]. They found relationships between variations in the gene MC1R and measured hair colour (most strongly with b*), as well as with chemical studies. It should be noted that people with red hair were deliberately over-represented in this sample and volunteers ranged in age from 6 to 72 years old (median of 35). Both studies emphasised the necessity and demonstrated the feasibility of measuring pigmentation objectively and accurately.

3.1.3 Aims

In every investigation, 'hair colour' must first be defined by the researchers. What one investigation calls 'blonde' however, might be what another calls 'light brown' and this inhibits the reproducibility of experiments and the direct comparison of results.

The aim of the research presented here was to objectively measure human hair colour and examine the variation found in a population with European ancestry, over all three colour axes and to analyse the variation of the measured hair colours to determine if and how hair colours form describable groups and how these match with our intuitive descriptions of hair colour.

3.2 Materials and Methods

3.2.1 Subjects

Subject recruitment and sampling procedures were conducted with the approval of the Victoria University Human Research Ethics Committee (approval number HRETH06/156). In total, 202 subjects were included in this analysis. All volunteers recruited for the study had their natural hair colour at the time of sampling. Most volunteers were between the ages of 18 and 35 and were of European ancestry (194 subjects). The exceptions to these criteria include six individuals of non-European ancestry (one African and five South Asians) and two older European individuals with white hair. Subjects fitting these criteria were chosen to meet the study aims of examining natural variation in adult European hair colour, with non-European and mature-white haired individuals included for comparison purposes.

3.2.2 Sampling

Subjects were first given all project information and the following procedures carried out following their written consent. Subjects filled out a questionnaire that, in addition to confirming their age, ancestry and natural hair colour status, asked them to report how they saw their own hair and eye colour (eye colour data was not analysed for this project). Hair colour was reported to be one of; Black, Dark Brown, Light Brown, Blonde, Red or White. Colour assessment by an observer (the researcher taking the samples) was also recorded. The project information sheet, consent forms and questionnaire can be found in Appendix 3.

Macroscopic hair colour was measured by reflective spectrophotometry (RS) using a Minolta CR-300 Chroma Meter (Konica Minolta, North Ryde NSW, Australia). The instrument was calibrated using a white tile (supplied by the manufacturer) and a light source input setting of 'D65' which represents daylight without spectral highlights. The instrument was set to measure in the CIE L*a*b* format. Hair colour was measured on the left, right and back of the head and was measured five times at each spot (See Figure 3.7), the mean of the 15 measurements being the focus of subsequent statistical considerations.

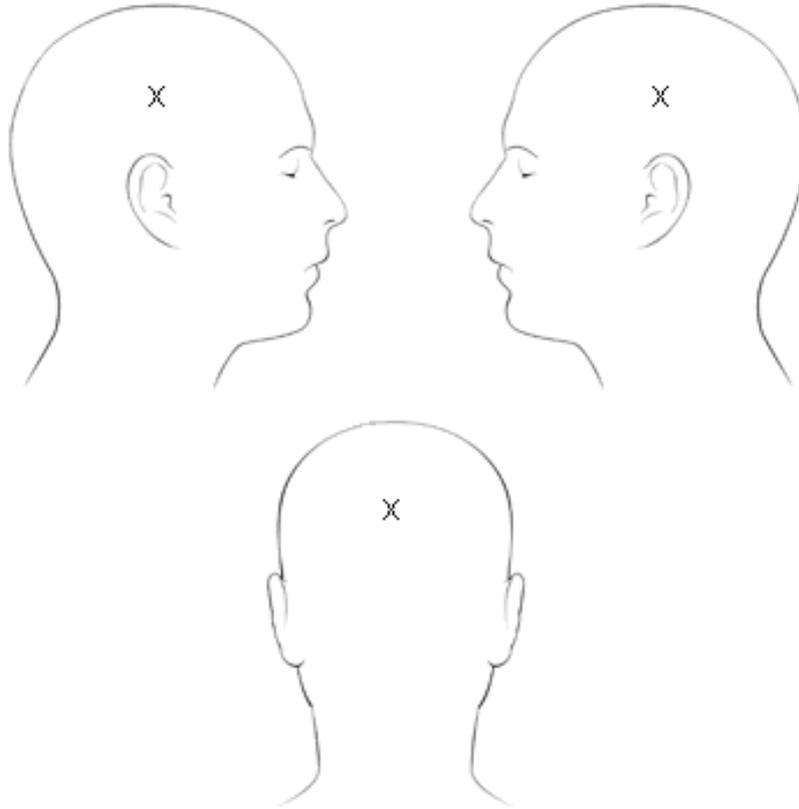


Figure 3.7 The locations on the head where reflective spectrophotometry measurements were taken; 5 measurements were taken at each site.

3.2.3 Statistics

SPSS 15.0 for Windows (© SPSS Inc. 2006, Chicago IL) was used to analyze the colour data. Two separate methods were used for grouping the population; a cluster analysis and discriminant analysis. Measured hair colour (by RS) was broken into several groups using a two-step cluster analysis; performed using the Euclidian distance criterion or the log-likelihood probability approach and either not specifying the number of clusters to be formed or asking for a specific number of clusters, ranging from two to seven. Cluster analysis seeks to identify natural subgroups within a population by minimising within-group variation while maximising between-group variation. To characterise and evaluate the clusters a separate, discriminant analysis was then performed. Discriminant analysis commences from the perspective of there being a known number of population sub-groups. From the population, there are a number of individuals whose sub-group classification is known. The data from these individuals is analysed in an attempt to build a profile of sub-group membership and subsequently use this profile to classify new individuals (whose sub-group membership is otherwise unknown).

After the analysis, SPSS determines which variables are most important in discriminating the groups, and how good this model will be at predicting group membership of future individuals. It determines the predictive value of the model by removing each individual separately, reanalysing the data and then predicting the membership of that case. The percent correctly classified is reported.

3.3 Results

3.3.1 Reported Colours

Self-reported hair colours (SRC) were compared to the colours determined by the observer (observer-reported colour or ORC) (Table 3.1). 85.2% of individuals had the same hair colour reported by themselves and the observer. Where there was a difference, discrepancies were lighter or darker by one shade. The observer was more likely (53.6% of disagreeing observations) to see a darker shade. In one case, a self-reported Red colour was reported as Blonde by the observer.

Table 3.1 The self-reported hair colours (SRC) compared to the observer-reported hair colours (ORC) of individuals with European ancestry.

Colour	Self Reported	Observer Reported
Black	4	10
Dark Brown	85	71
Light Brown	50	62
Blonde	41	38
Red	14	13
White	2	2

3.3.2 Population Variation

To examine the hair colour variation amongst the European population, the L^* , a^* and b^* values of the 196 European individuals, as measured by reflective spectrophotometry, were recorded and the data subjected to statistical analysis. The variability observed in the entire sample is illustrated in Figure 3.8. This distinctiveness of the non-European individuals (which are included and labelled as NE in Figure 3.8) and the individuals with white hair (W in Figure 3.8) can be seen in the highlighted areas. Individuals not of European ancestry have hair that is much darker with a^* and b^* values very close to zero.

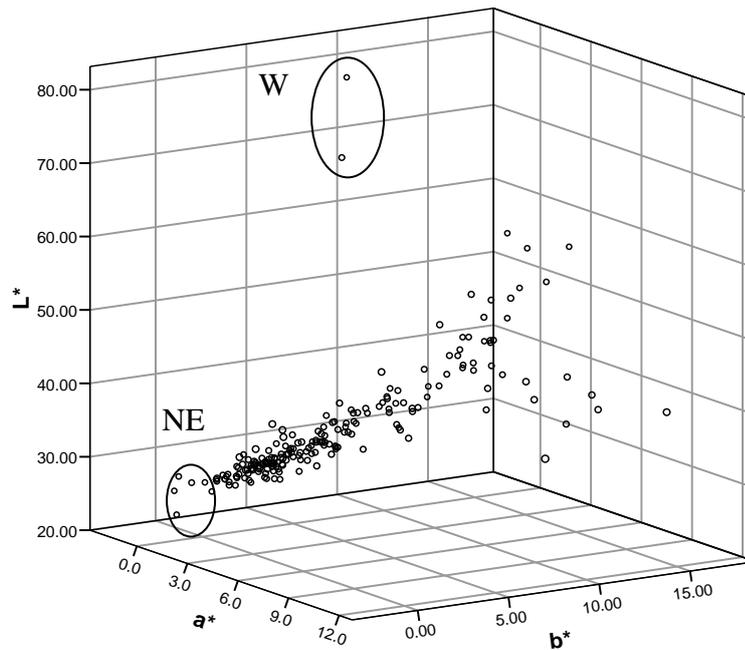


Figure 3.8 Population Sample; the non-European samples (NE) and the mature-white haired samples (W) are circled. The scales of the CIE L*a*b* colour space are; L* = 0(black) to 100 (white), a* = -100 (green) to +100 (red) and b* = -100 (blue) to +100 (yellow). Only the relevant region of the scale is shown.

3.3.3 Grouping the Population

The aim of the statistical analysis was to investigate the major components of the colour variation and to see whether hair colour falls into groups that would be meaningful for further research. What qualifies as a meaningful group is subjective and may depend on the intended use of the data; however in this case, meaningful groups would be those that separate colours into several distinct groups. For practical purposes, it would also be useful to have groupings that correspond well to colours people already use to report hair colour (e.g. Blonde).

The European sample was broken into groups using the Two-Step cluster analysis. The analysis was first conducted without specifying the number of clusters, using either the Euclidian distance approach (E) or the log-likelihood probability approach (LL). Next, the numbers of clusters were specified, from two to six for both approaches (LL2 to LL6 and E2 to E6) and a discriminant analysis was then performed. Using the clusters determined as a definitive population sub-group, the percentage of cases correctly classified was used to evaluate the clusters. A maximum of six clusters were analysed since there were six observer-reported categories and since discriminant scores decreased continuously with more clusters (data not shown).

The percentage of cases correctly classified in the discriminant analysis for the various methods of cluster determination is shown in Figure 3.9. When the observer-reported colour (ORC) groupings are used, only 70.4% of cases are correctly classified. Almost all other clusters have greater than 95% of cases correctly classified. Only E4 had less than 95.0% but was very close at 94.9%.

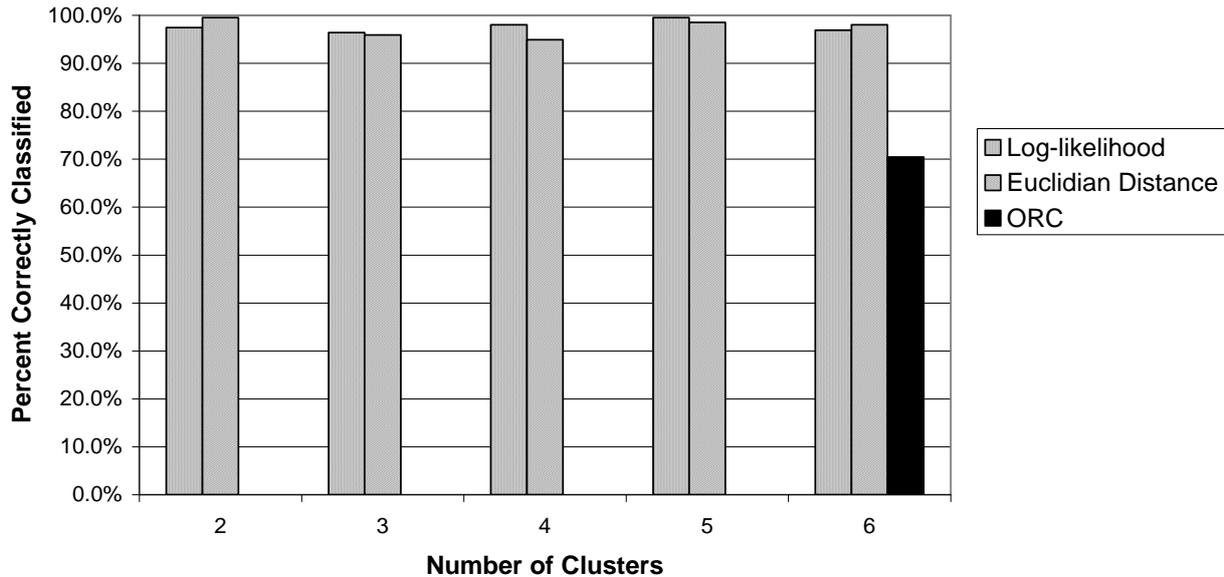


Figure 3.9 Discriminant scores for each number of clusters and the method by which they were determined.

Using the Euclidian distance measure to determine clusters did not result in as useful groupings. In all cases, between 67% (for six clusters) and 93.9% (for two clusters) of the sample was in one cluster (data not shown). This is a problem because it results in one large cluster with only extreme individuals in several very small clusters.

The log-likelihood probability method of clustering produced several promising groupings (Table 3.2 to Table 3.4). The best result for the purposes of this study was obtained with five clusters (LL5). This has the highest discriminant score (99.5%) and the most appealing cluster characteristics (Table 3.2) which will be discussed later in more detail. Other numbers of clusters determined by log likelihood may be useful to different types of studies and have good discriminant values. For example, six clusters (LL6) would be useful when comparing to the six observer-reported groups discussed in Section 3.3.4 and in Figure 3.11, while two clusters (LL2) would be useful for dividing

a population into light and dark hair colours. Both of these analyses have good discriminant scores (97.4% and 96.9%, respectively) and good group composition.

Table 3.2 Composition and cluster characteristics of five clusters determined by the log-likelihood probability method (LL5)

	N	% of Total	L*		a*		b*	
			Mean	SD	Mean	SD	Mean	SD
Cluster 1	8	4.1%	39.58	2.74	8.29	1.39	12.79	1.95
2	32	16.3%	46.80	5.38	4.42	0.65	11.19	2.01
3	2	1.0%	74.62	6.27	0.26	1.24	7.57	1.34
4	68	34.7%	35.46	2.44	3.22	0.62	4.82	2.26
5	86	43.9%	30.41	1.63	1.70	0.63	1.80	1.19
Total	196	100%						

Table 3.3 Composition and cluster characteristics of two clusters determined by the log-likelihood probability method (LL2)

	N	% of Total	L*		a*		b*	
			Mean	SD	Mean	SD	Mean	SD
Cluster 1	47	24.0%	45.88	8.40	4.87	1.96	10.99	2.33
2	149	76.0%	32.44	3.08	2.31	0.93	2.96	2.13
Total	196	100%						

Table 3.4 Composition and cluster characteristics of six clusters determined by the log-likelihood probability method (LL6)

	N	% of Total	L*		a*		b*	
			Mean	SD	Mean	SD	Mean	SD
Cluster 1	8	4.1%	39.58	2.74	8.29	1.39	12.79	1.95
2	27	13.8%	47.79	5.19	4.41	0.65	11.65	1.66
3	2	1.0%	74.62	6.27	0.26	1.24	7.57	1.34
4	27	13.8%	38.48	2.24	3.79	0.72	7.51	1.30
5	75	38.3%	30.14	1.54	1.58	0.55	1.73	1.12
6	57	29.1%	33.92	1.98	2.92	0.52	3.39	1.78
Total	196	100%						

During analysis, canonical discriminant functions are determined. These unstandardised functions are used to make classifications in the discriminant analysis and the standardised function coefficients are used to compare the relative importance of each function and each variable in making classifications. For LL5, three discriminant functions were determined; $F1 = -7.359 + 0.089L^* + 0.953a^* + 0.518b^*$ which contributes 66.6% of the information for discriminating clusters, $F2 = -9.759 + 0.373L^* - 0.822a^* - 0.233b^*$ with 32.3% of the information and $F3 = -4.802 + 0.140L^* + 0.826a^* - 0.534b^*$ with 1.1% of the information. In Figure 3.10, the results of the first two functions are used to illustrate the clusters.

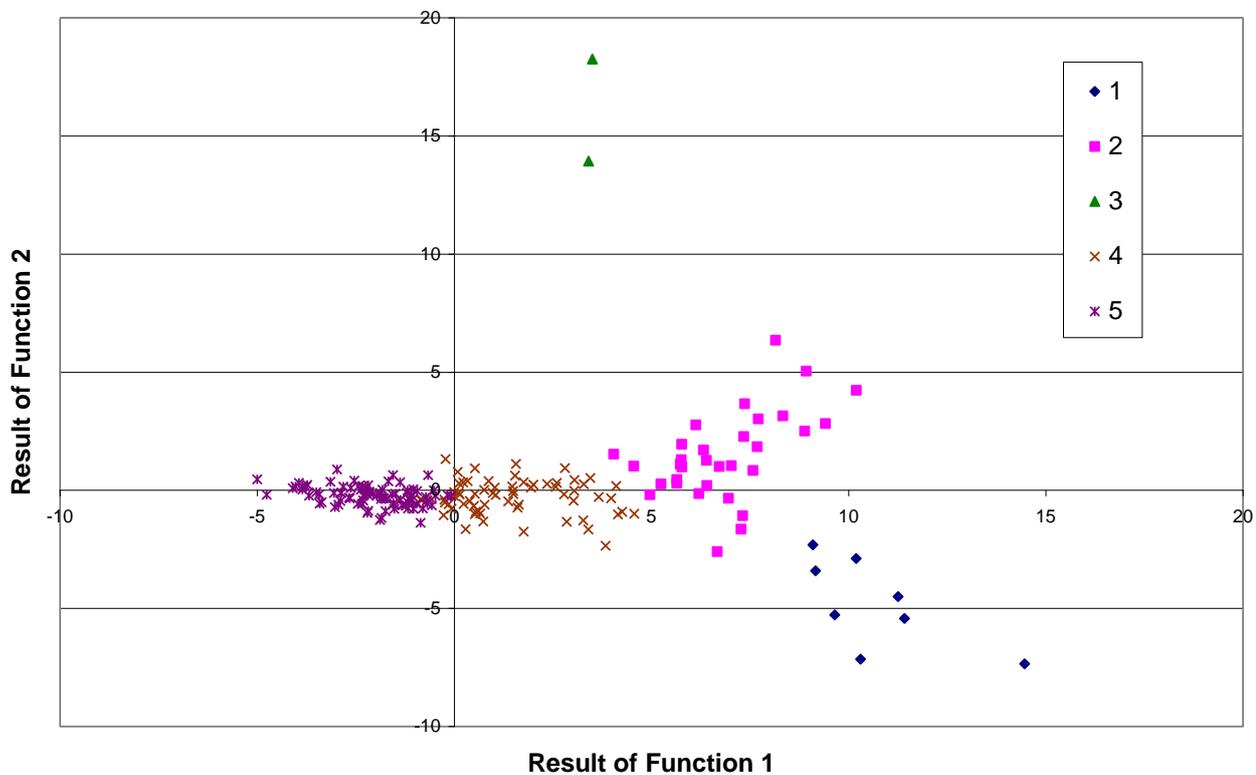


Figure 3.10 Five clusters determined by the log-likelihood probability method (LL5) separated according to the results of the first two significant discriminant functions. $F1 = -7.359 + 0.089L^* + 0.953a^* + 0.518b^*$ and $F2 = -9.759 + 0.373L^* - 0.822a^* - 0.233b^*$

3.3.4 Comparing Clusters to the Observer-Reported Colours

It appears that the clusters determined by various methods show similarity to hair colours as perceived by an observer. Looking at the LL5 clusters in Figure 3.10 again, the clusters could be referred to as Red, Fair, White, Medium and Dark (for clusters 1 to 5, respectively) by both their $L^*a^*b^*$ values (Table 3.2) and the observer reported colours they contain (Table 3.5). The names are arbitrary; however they correspond to how people tend to refer to hair colour.

Table 3.5 Observer reported colours found in the LL5 clusters and descriptions of individuals compared to other ORC group members.

LL5 Cluster	Possible Cluster Name	Observer Reported Colours Present
C1	Red	8 Red (highest a^* and b^* values)
C2	Fair	29 Blonde 3 Red (less a^* and b^* than C1)
C3	White	2 White
C4	Medium	10 Blonde (darker and less a^* and b^* than C2) 41 Light Brown 15 Dark Brown (lighter and higher a^* and b^* values than C5) 2 Red (less a^* and b^* than C1, darker than C2)
C5	Dark	16 Light Brown (darker and less a^* and b^* than C4) 62 Dark Brown 8 Black

While the five clusters determined in LL5 are the best for the purposes of this project, it is worthwhile illustrating a direct comparison between the six observer-reported categories and six clusters produced by SPSS. Figure 3.11A shows the six clusters as determined by their reflective spectrophotometric measurements (LL6), separated according to the discriminant functions while Figure 3.11B shows the same individuals grouped into their observer-reported categories. The lack of separation using observer-reported categories is obvious and reflects the low mathematical discrimination ability of these categories. It should be noted that clusters are numbered independently in each analysis; Cluster 1 in LL5 does not necessarily contain any or all of the same individuals as Cluster 1 in LL6.

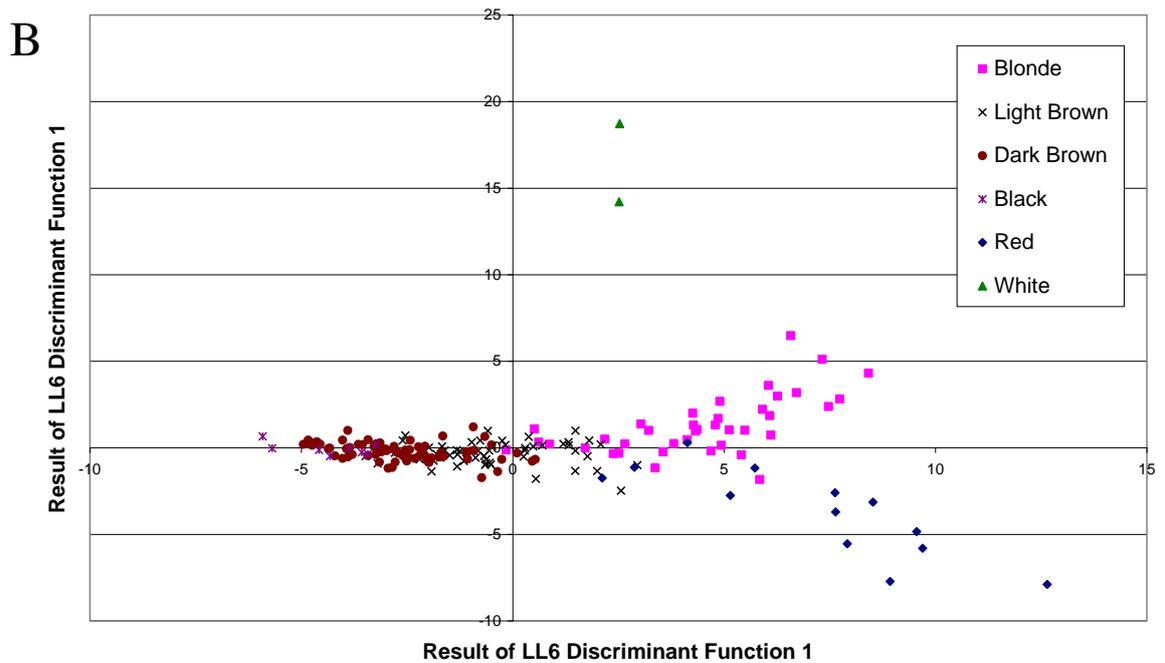
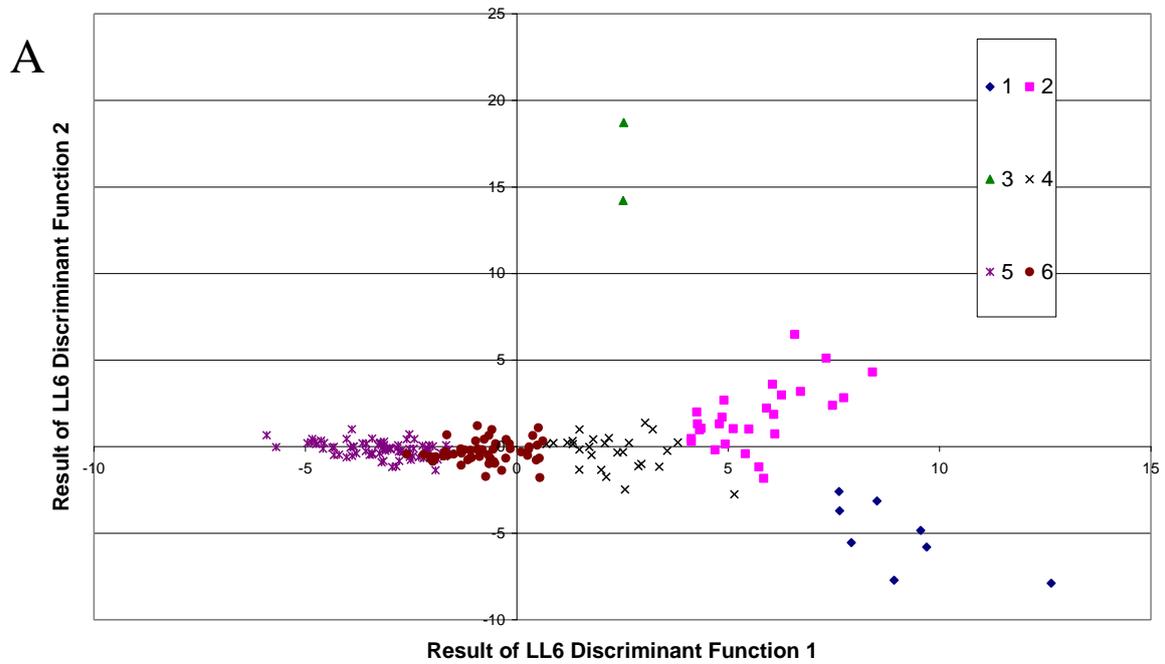


Figure 3.11 The two most important functions used to discriminate the six clusters as determined by log-likelihood (LL6) were applied to the $L^*a^*b^*$ data and grouped by Cluster (A) and by observer-reported colour (B). $F1 = -8.558 + 0.101L^* + 0.972a^* + 0.435 b^*$ and $F2 = -9.671 + 0.376L^* - 0.908a^* - 0.222b^*$

3.4 Discussion

The presented data show that when given a limited choice of colours, self-reported hair colours and those reported by an observer are fairly consistent; however, these categories are not necessarily the best way to categorise hair colour for quantitative research. While hair colour can be perceived as a continuum (as seen in Figure 3.8), it can be divided into categories or clusters based on spectrophotometric measurements in the CIE $L^*a^*b^*$ colour space and these clusters can be well discriminated from each other. The discrepancies between self-reported hair colour and observer-reported hair colour and the poor separation of observer-reported colour groups compared to those defined by clustering analysis emphasise the importance for objective measurement of individuals to be included in research studies. For repeatability and validity of studies, phenotypes must be defined as well as possible and the objective measurement methods and analysis strategies presented here may be of assistance in this regard. The reduction of three colour components to one or two discriminant functions makes the examination and classification of individuals easier to represent.

It is also interesting to note that, while not exact, clustered groupings follow the same pattern that self-and observer reported colours do. How someone decides to report a hair colour may be an artefact of the arbitrary naming of colours or of personal bias or it may also be due to human perception of colour, which corresponds to the CIE $L^*a^*b^*$ system of measurement or it may be all of the above. The human eye is a complicated structure with retinal receptors (rods and cones) that are sensitive to changes in lightness and colour [55] as discussed in Section 3.1.1. This system of human trichromatic vision may be why similar groups are defined by both the human reporting of colour and by the spectrophotometry and clustering method described here. The lack of sensitivity in biological perception may contribute to the variability found. As mentioned previously, one unit on the $L^*a^*b^*$ scale is defined as the limit of human discrimination between colours [57]. The difference between members of different clusters can be less than this in one or more of the three colour axes and may contribute to the inconsistency between reported colour groups and clusters determined by spectrophotometric data and clustering algorithms.

The particular clustering algorithms and program options used will depend on the intended use of the data. For example, a medical genetic association study may require

a Fair/Dark division, where the prediction of hair colour in a forensic study may require more specific groupings. This study has illustrated the possibilities by using a two-step cluster analysis in SPSS; however there are many other programs and clustering algorithms available. Having objectively measured colours allows researchers to group individuals in clearly defined ways and to change these groups as the study demands without subjectivity or collecting additional data.

It is acknowledged that the exact functions and cluster details may change as larger samples are analysed. In this study, the composition of clusters changed when the sample size was increased from 134 to 196 (See Vaughn *et al.* 2008 in Appendix 4). LL2 and LL6 retained their high discriminant scores and good group composition; however E4 did not retain its high discriminant score and was replaced by LL5. It is expected that a sample size would eventually become sufficiently large to resist major changes. Using smaller sample sizes with colour measurement data, the researcher has the options of changing clusters for each sample-set or defining groups based on an initial sample-set and assigning new individuals according to the associated discriminant functions.

As mentioned, different clustering results may be appropriate for different types of studies. For future considerations in this research, the LL5 clusters are the most appealing. This grouping conveniently divides the non-red (blonde-black) spectrum into three large groups and recognises those with Red and White hair as separate. This corresponds nicely with the study aims of studying non-red colours.

Reflective spectrophotometry, while accurate, may be inconvenient for measurements on very large numbers of people or for investigators. Chapter 4 will investigate the use of digital images for measuring and classifying colour.

Chapter 4 A comparison of Hair Colour Measurement by Digital Image Analysis with Reflective Spectrophotometry³

4.1 Introduction

4.1.1 Digital Photography and Digital Colour Measurement

Traditional photography captures images by exposing light sensitive silver compounds while digital photography captures images by a charge coupled device (CCD) array which transforms the absorbed light energy into a digital image [62] that can be viewed on a computer screen. Figure 4.1 illustrates the two processes of image production.

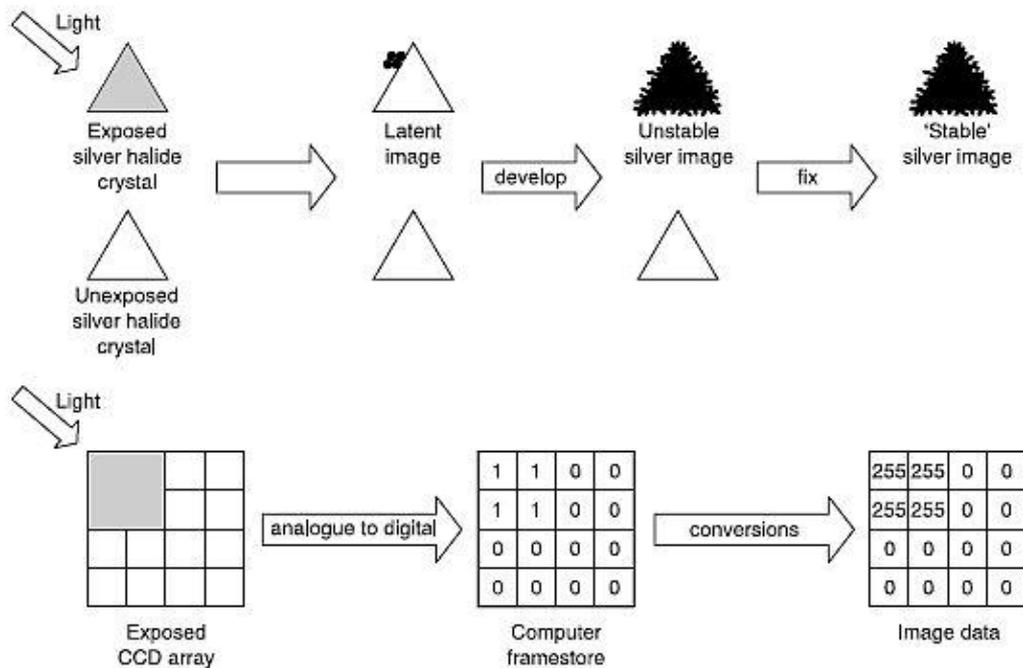


Figure 4.1 Image formation by silver halide and CCD sensors [62]

The reproduction of colour in photography is a complex issue and while no reproduction of colour is identical with the object being photographed, modern imaging

³ An earlier version of this experiment was published in *Forensic Science International*. Vaughn *et al.* 2009:183 Pages 97-101. Please see Appendix 4

systems are able to reproduce colour very well [62]. Digital images are usually displayed in the RGB colour space (Discussed in Section 3.1.2) on a computer monitor or similar device. The RGB colour system is composed of three values (Red, Green and Blue) that are measured from 0 to 255. Therefore a colour 0,0,0 is pure black while 255, 255, 255 is pure white and various colours or hues are made of combinations of red, green and blue [57]. Figure 4.2 illustrates the colours and related values of the RGB model.

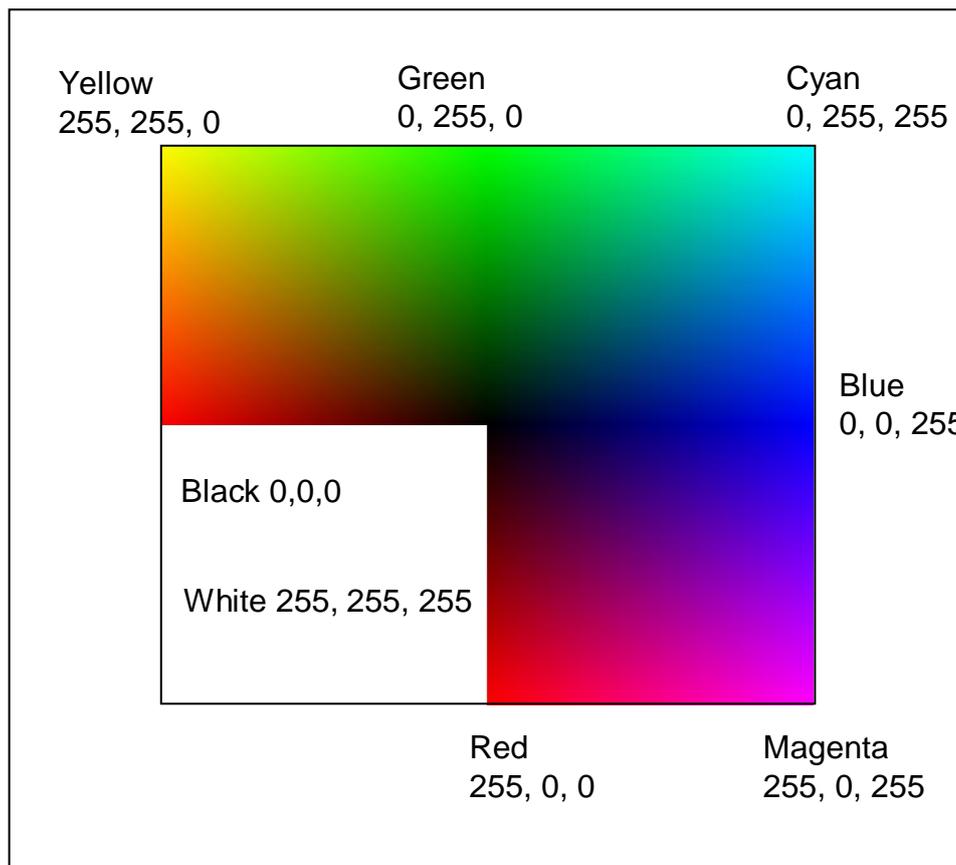


Figure 4.2 Colours and values within the RGB model

Colours can be measured by various software applications by reporting the intensities of the red, green and blue in the image. Conversion to other colour spaces is then a matter of mathematical algorithms [63]. Different colour spaces are not equivalent to one another but images can be measured in various spaces or the values converted. For this experiment, we are concerned with converting RGB to the CIE L*a*b* space discussed in 3.1.2. The formulas for converting RGB to L*a*b* from Herbin *et al.* [63] are:

$$\begin{aligned} L^* &= 25 * (100 * Y/Y_0)^{1/3} - 16 \\ a^* &= 500 * [(X/X_0)^{1/3} - (Y/Y_0)^{1/3}] \\ b^* &= 200 * [(Y/Y_0)^{1/3} - (Z/Z_0)^{1/3}] \end{aligned}$$

where XYZ is the CIE system defined by:

$$\begin{aligned} X &= 0.618 * R + 0.177 * G + 0.205 * B \\ Y &= 0.299 * R + 0.587 * G + 0.114 * B \\ Z &= 0.000 * R + 0.056 * G + 0.944 * B \end{aligned}$$

and X_0, Y_0, Z_0 , are the values of a specific white reference.

While reflective spectrophotometry is an established method for measuring hair colour at the macroscopic level, it can be cumbersome to use on a large number of individuals. This study investigates the use of digital photographs to measure hair colour and compares its use to the reflective spectrophotometry standard. An understanding of the accuracy of the colour determination by these methods is of relevance when undertaking specific investigations, such as those on the genetics of hair colour. Measurement of hair colour by digital image analysis may also be of assistance in cases where a photograph is the only evidence of hair colour available (e.g. surveillance).

Digital image analysis has previously been used to measure hair colour in forensic science; however it has generally been used to measure the colour during the microscopic examination of single hairs and not the macroscopic colour of hair on the head [24]. This has been done to assist the traditional forensic hair examination, with the purpose of attributing a hair to a specific donor. Digital images have been used extensively to measure the colour of many other objects, such as vegetables, soil and chocolate [64-66], and have been used to measure skin colour [63, 67, 68]. Herbin *et al* [63]. found that the colour of skin lesions measured from 12 digitised images of colour slides was highly correlated with values obtained by reflective spectrophotometry, with correlation coefficients of 0.97, 0.99 and 0.98 for L^* , a^* and b^* of the CIE $L^*a^*b^*$ model (see Figure 3.5) respectively. Takiwaki *et al.* [67] achieved similarly good results with eight videomicroscopic images of skin.

When measuring colour from digital images, it is important to note that digital images are most often displayed and measured in the Red-Green-Blue (RGB) colour space of the computer monitor. There are various programs that may be used to measure this

colour and then to convert it into the colour space required (CIE L*a*b* in this case) using standard mathematical algorithms [63]. This study considers two such programs.

4.1.2 Aims

The aim of this study is to compare the results obtained from measuring the colour from digital images (DI) of hair colour (at the macroscopic level) to the measurements of the same individuals using reflective spectrophotometry (RS) as a standard, to see how the measurements correspond. This will help determine to what extent digital image analysis can be used instead of reflective spectrophotometry and, if so, in what situations.

4.2 Materials and Methods

4.2.1 Subjects

Described in 3.2.1

4.2.2 Sampling

In addition to the reported colours and reflective spectrophotometry measurements described in Section 3.2.2, digital photographs were taken of the scalp hair. Digital images were taken of the left, right and back of the head with a CASIO 4.0 mega pixel camera (model QV-R40) with an image size of 640x480 pixels and with picture quality set to 'fine'. On 'collection' mode (for taking close-up photographs), the flash was turned off, the white balance tool was set to 'sunlight' and all other settings were left at the default. Each time a set of images was to be taken under a different light source (on a different day, different time of day, different room, etc.) a piece of new white paper was photographed in the same position as the head of the subject (three sides) and measured using the Minolta CR-300. This 'white image' was later used to correct colour for light source. All images were taken at a distance of 15cm in the same locations as the reflective spectrophotometry readings (See Figure 4.3 and compare to Figure 3.7).



Figure 4.3 Sample photographs before cropping and analysis

4.2.3 Digital Image Preparation and Analysis

Software packages used included:

- JASC Paint Shop Photo Album 4.0 by Jasc Inc.
- Adobe Photoshop Version 9.0.2 © 1990-2005 Adobe Systems Inc.
- IrfanView 3.92 © 1996-2004 Irfan Skijan; Vienna University of Technology
- MATLAB 4.0 © 1994-2006 MathWorks with the Image Processing Toolbox and modules written to convert RGB colour measurements to $L^*a^*b^*$ measurements [64]
- V++ Version 4.0 ©1990-2002 Digital Optics Ltd. New Zealand, with modules written to convert RGB colour measurements to $L^*a^*b^*$ measurements [64, 69]
- ImageJ ©1997-2006 by W.S. Rasband at the US National Institutes of Health, with the RGB-to-Lab conversion plugin installed [70].

Each digital image of the hair was cropped to 100x100 pixels in the area closest to the centre of the image (where the RS measurements were taken) with JASC Photo Album, Adobe Photoshop or IrfanView as available. Care was taken to avoid obvious shadows or spectral highlights that result from the curvature of the skull.

Initially, images were analysed using MATLAB to convert the RGB colour measurements to $L^*a^*b^*$ and to determine the mean colour values for each image; however, this program is rather difficult and tedious for a new user. The RGB to $L^*a^*b^*$ conversion modules written for MATLAB were instead used within V++ [69], a more user-friendly program, which results in identical values as those determined by MATLAB. Therefore, V++ was used to analyse all subsequent images. The computer code for the RGB2LAB conversion module can be found in Appendix 2.

For each image, the L^* , a^* and b^* values were calculated. While a reflective spectrophotometer has its own, consistent light source and can be calibrated to a white tile (as described in Section 3.2.2), a digital camera cannot be calibrated before the photograph is taken and since hair is shiny and reflects a flash, no consistent light

source was available. Images of the same colour taken under different light sources will appear different [63], so an effort was made to correct this colour in a manner similar to a previous study using the RGB colour space [63], with the understanding that perfect correction is unlikely and that the greatest difficulty in correcting the colours (a^* and b^* axes) will be found when the lightness of the image differs. The new white paper, mentioned above, when measured with the Minolta Chroma Meter CR-300 was found to have L^* , a^* and b^* values of 93.4, 1.16 and -2.06 respectively. Using this as a reference point, L^* , a^* and b^* colour points were added or subtracted depending on how much the white reference paper deviated from its Minolta CR-300 determined values under the various sampling situations. For example, if L_i is the L value of the colour image, L_w is the L value of the white image and L_c is the L value of the corrected image, then: $L_c = (93.4 - L_w) + L_i$.

4.2.4 Correction

To examine the accuracy of the correction formula, images were taken of white, pink and blue paper under four different light sources i) fluorescent light ii) fluorescent light and natural sunlight iii) incandescent light and natural sunlight and iv) bright natural sunlight. The images of the paper were cropped to 100x100 pixels and analysed for colour with V++. The pink and blue colours were corrected using the white values obtained from V++, using the above formula and compared to the RS measurements (true colour) of the paper.

4.2.5 Program Comparison

To examine whether another program is better able to match the RS values through its algorithm or data handling, the MATLAB/V++ module for RGB to $L^*a^*b^*$ colour conversion was compared to that of the ImageJ program. Values were calculated for the pink, blue and white paper images and corrected using the white paper values calculated from Image J and with the above formula. The corrected ImageJ values were compared to the RS measurements and to the V++ values.

4.2.6 Statistics

SPSS 15.0 for Windows (© SPSS Inc. 2006, Chicago IL) was used to analyze the colour data. Basic statistical tests were performed, including a One-Sample T-test (two tailed) for a difference in mean, a Paired T-test (two tailed) and an F-test for a difference in variance [71]. For details on the cluster analysis using reflective spectrophotometric data, please see Section 3.2.3 [72].

Delta E* (ΔE^*) is an Euclidian measure of the total distance between two CIE L*a*b* colour values and is calculated according to the following formula [55]:

$$\Delta E^* = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}$$

4.3 Results

4.3.1 Colour Correction

When comparing the colour mean values of pink paper measured by digital image analysis to the reflective spectrophotometric value, it was found that the DI a^* component values have means that are significantly lower than the RS measurements, whether the correction formula is used or not. Using a One sample T-test, there was a 95% confidence interval of -1.51 to -5.96 for the difference between the uncorrected DI measurement for a^* and the RS measurement and from -0.89 to -8.27 for the difference between the corrected DI measurement for a^* and the RS measurement, thus neither of these CI ranges included zero, indicating a significant difference. There was no significant difference between the RS and DI means for the L^* or b^* values. For the blue paper, only the corrected L^* value was significantly different from the RS measurements with a 95% confidence interval (One sample T-test) of 3.78 to 9.90.

Although not all of the means were significantly different, it was noted that for the L^* and b^* values of Pink and Blue paper (four out of six measurements) the confidence intervals for corrected digital measurements were narrower than those for the uncorrected formulas, and so an F-test for the equality of variance was conducted (between the uncorrected and corrected values for pink and blue paper). The results indicate that the variance for L^* of pink paper and the L^* and b^* components of blue paper are significantly smaller using the correction formulas (F-test for variance, p -value < 0.05). It was decided to use the correction formula for further experiments due to the closer means (though not significant) and the significantly smaller variances.

4.3.2 Program Comparison

To test whether the colours obtained by the programs V++ and Image J are any different, a Paired T-test of the means was used for the corrected L^* , a^* and b^* values from each program. The only significant difference (p value = 0.002) was that the a^* value for the blue paper was slightly higher in ImageJ (-5.19) than in V++ (-5.99).

The variance of the measurements as determined by the two programs was also compared. There was no significant difference (F-test for variance; p -values > 0.1) between the variance of colour measurements as determined by V++ and ImageJ. It can

be concluded that the L*a*b* colour measurement functions of ImageJ and V++ can be used interchangeably.

4.3.3 Intra-individual Comparison

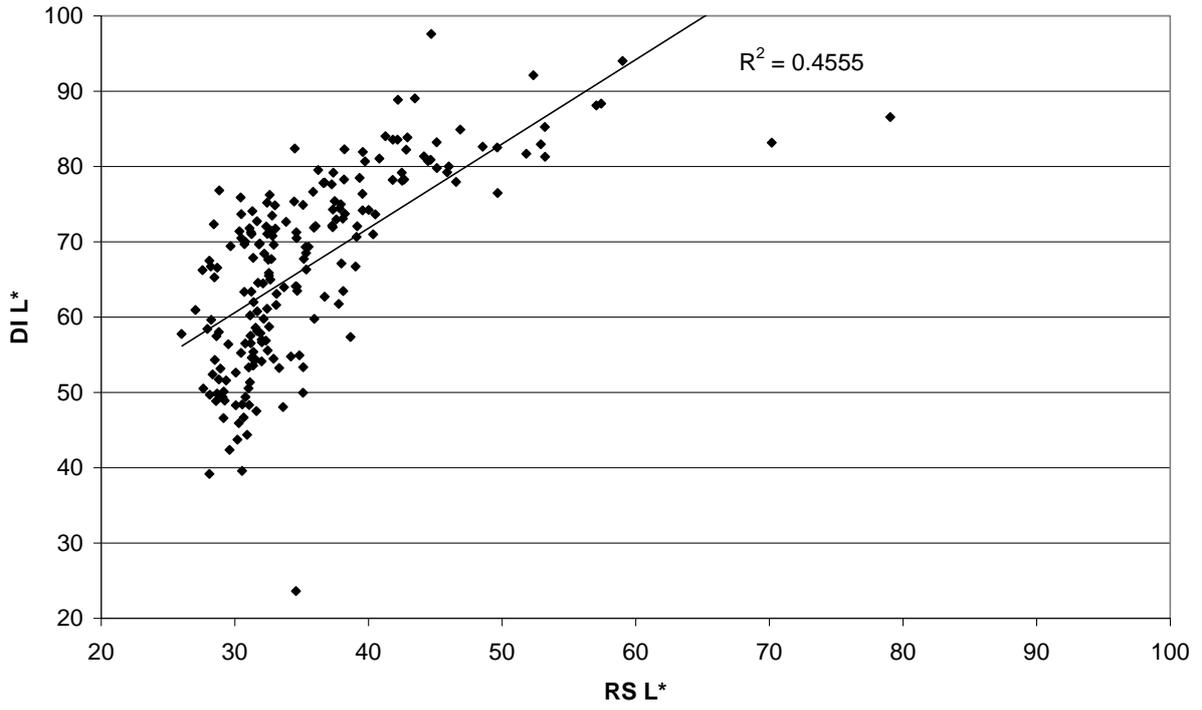
To test the consistency of the digital image colour measurements, 10 sets of digital images were taken of one individual (three photos for each set taken of the right, left and back of the head as seen in Figure 3.7). The mean L*, a* and b* of each set (of three images) was calculated and a one-way ANOVA was used to determine that these means were not significantly different over the 10 sets. This indicates the low amount of variation found across multiple measurements taken of one individual under the same conditions.

4.3.4 Digital Image Comparison to Reflective Spectrophotometry

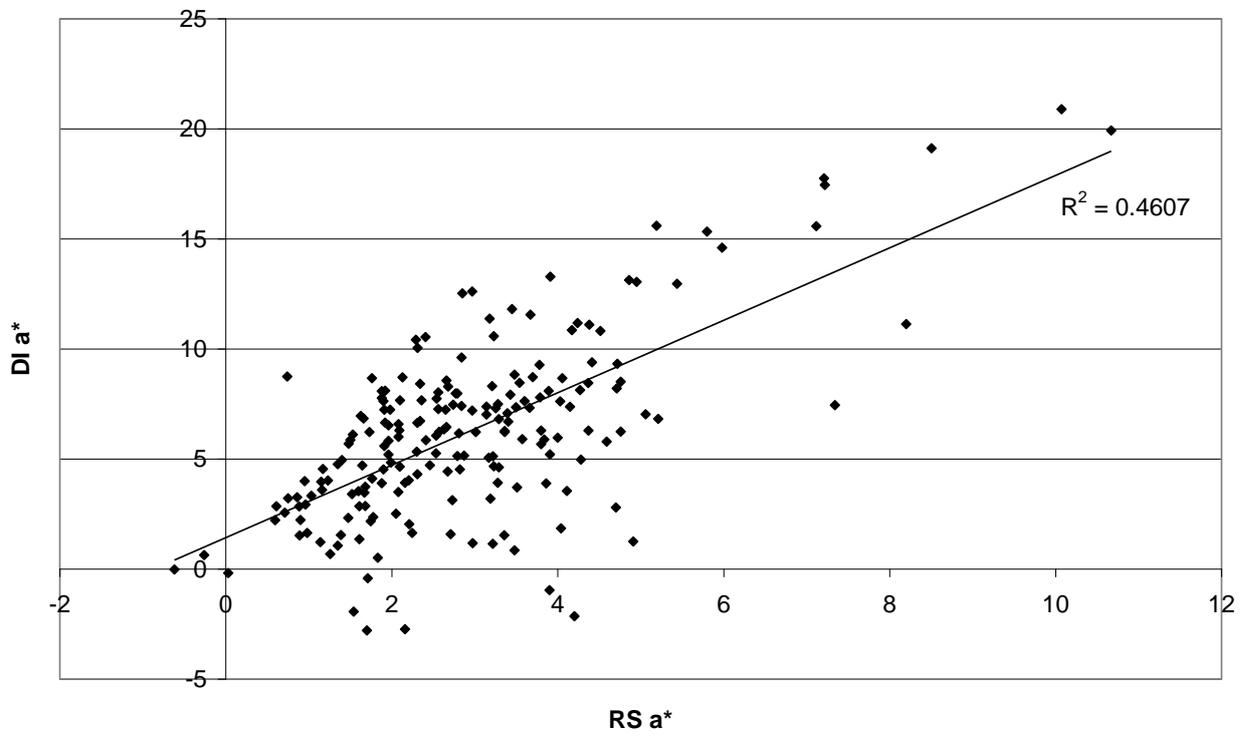
After determining that the correction formulas produce more consistent results and that two programs selected to analyse the digital images are almost the same, the entire European sample (196 individuals) was analysed to determine how digital image (DI) colour analysis compares to the colours as determined by reflective spectrophotometry (RS). The mean ΔE^* value for the difference between RS and DI colour was 34.16 units with a standard deviation of 9.10 units. However; this measurement does not give information about which components are different and therefore, each colour axis was examined separately.

Initially, the digital image colour measurements were compared directly to the RS measurements for the L*, a* and b* axes. A correlation was found along all three axes, with Pearson Correlation Coefficients of 0.675, 0.679 and 0.537 for L*, a* and b* respectively (p-values = 0.000). This is a significant correlation, but of only moderate strength. The coefficient of determination (R^2) indicates that only 45.6%, 46.1% and 28.8% of the variation (for L*, a* and b*, respectively) can be explained by the relationship between the RS and DI measurement. Figure 4.4 illustrates this relationship.

i)



ii)



iii)

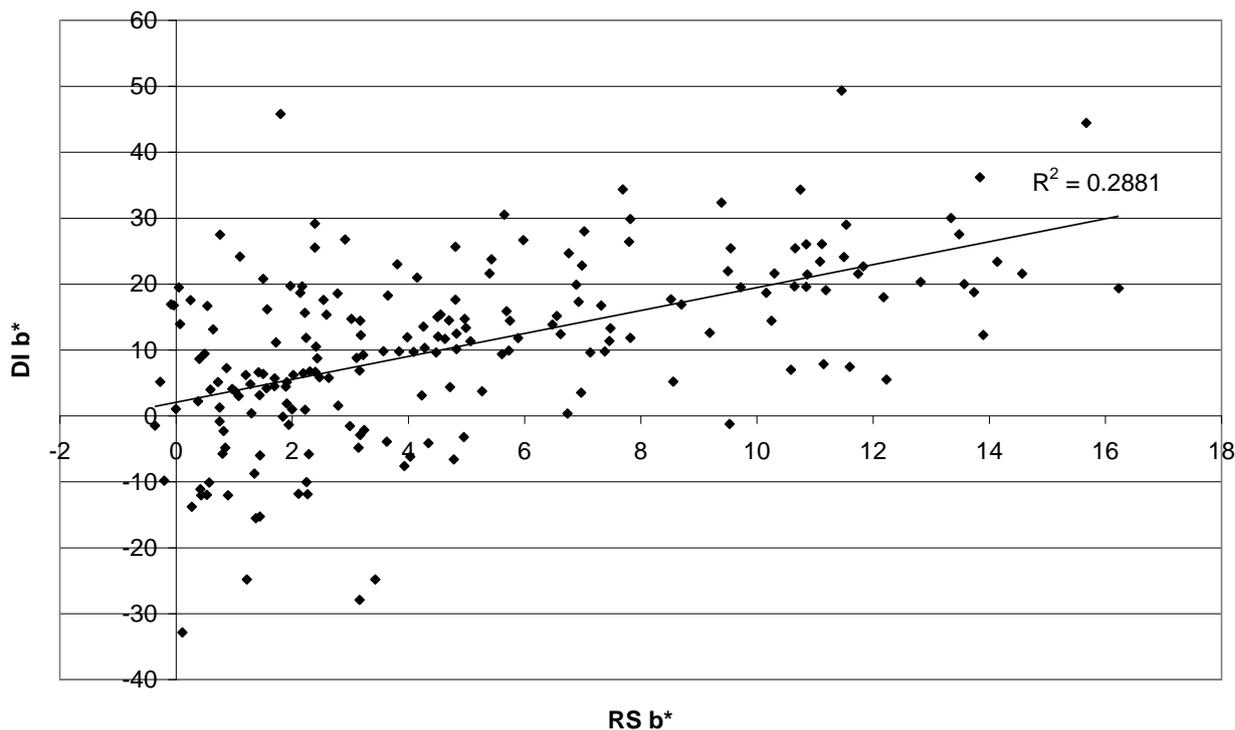


Figure 4.4 The relationship between RS and DI measurements for i) L* ii) a* and iii) b*

To see if there is a pattern in the differences between DI and RS measurements, a Paired T-test for the difference between the means obtained by DI and RS was performed. Means are significantly overestimated by digital image analysis for all three colour components. The L* value is significantly higher by an average of 31.28 units, the a* value higher by an average of 3.32 units and the b* value higher by an average of 5.68 units (p-values = 0.000).

To investigate whether the overestimation error was uniform across the sample, or if light or dark hair is more prone to overestimation, the difference between the DI and RS measurements was compared to the original colour (RS measurement). The difference in L* showed no significant correlation, indicating that the amount of overestimation is not related to the original hair colour for this component; however, there was a low amount of correlation (R = 0.341) between the overestimation of the DI a* and the RS a* values (p-value = 0.000) and between the overestimation of the DI b* and the RS b* values (R = 0.261, p-value = 0.000) indicating that there is a slight tendency for the

difference in a^* values to become greater as hair becomes more red and for the difference in b^* values to become greater as hair becomes more yellow.

As the separate colour components (L^* , a^* and b^*) measured by RS and DI do not correspond well with each other, it was investigated whether, when combined, they would place individuals into the same colour categories as those determined by RS measurements.

Chapter 3 showed that the RS measurements can be used to break this same sample set into defined groups using a Two-Step Cluster analysis. These groups were then evaluated using a discriminant analysis to classify cases into the correct groups. In Chapter 3, the best results were obtained for an analysis resulting in five clusters using a log-likelihood probability method (LL5) with a discriminant score of 99.5%. Other groupings may have some use, such as LL2 or LL6 and each of these obtained classification scores above 95%. This was better than the 70.4% discrimination obtained by using observer-reported clusters.

For the present study, each individual was assigned to a cluster based on their RS data (for each of the three analyses). These clusters were then re-evaluated, using their DI data and the percent of cases correctly classified is reported in Table 4.1. Table 4.1 also shows that the percent of cases correctly classified in the six observer-reported colour groups (52.0%) was lower than for any of the RS analyses.

Table 4.1 Percent of cases correctly classified by a discriminant analysis on Clusters determined by RS or the observer reported colour group (ORC), using DI data

Analysis	Number of Clusters	Percent Correctly Classified by Digital Image Analysis
LL2	2	82.1%
LL5	5	63.8%
LL6	6	58.2%
ORC	6	52.0%

Figure 4.5 illustrates an example of the use of discriminant functions in classifying individuals (using the results from LL5). The discriminant functions determined by RS data were $F1 = -7.359 + 0.089L^* + 0.953a^* + 0.518b^*$, $F2 = -9.759 + 0.373L^* - 0.822a^* - 0.233b^*$ and $F3 = -4.802 + 0.140L^* + 0.826a^* - 0.534b^*$. The first two functions were used to plot individuals onto a grid using first their RS measurements (Figure 4.4A) and then their DI measurements (Figure 4.4B). Figure 4.4A shows the good separation of the five clusters determined in LL5 while Figure 4.4B shows the groups merging into one another.

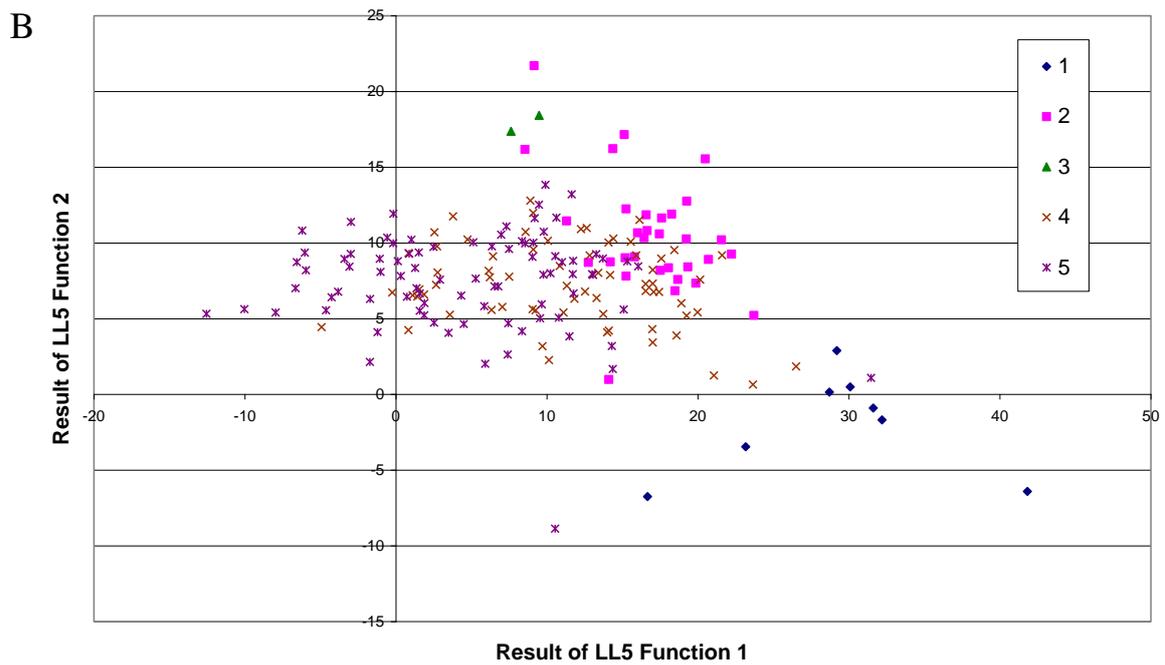
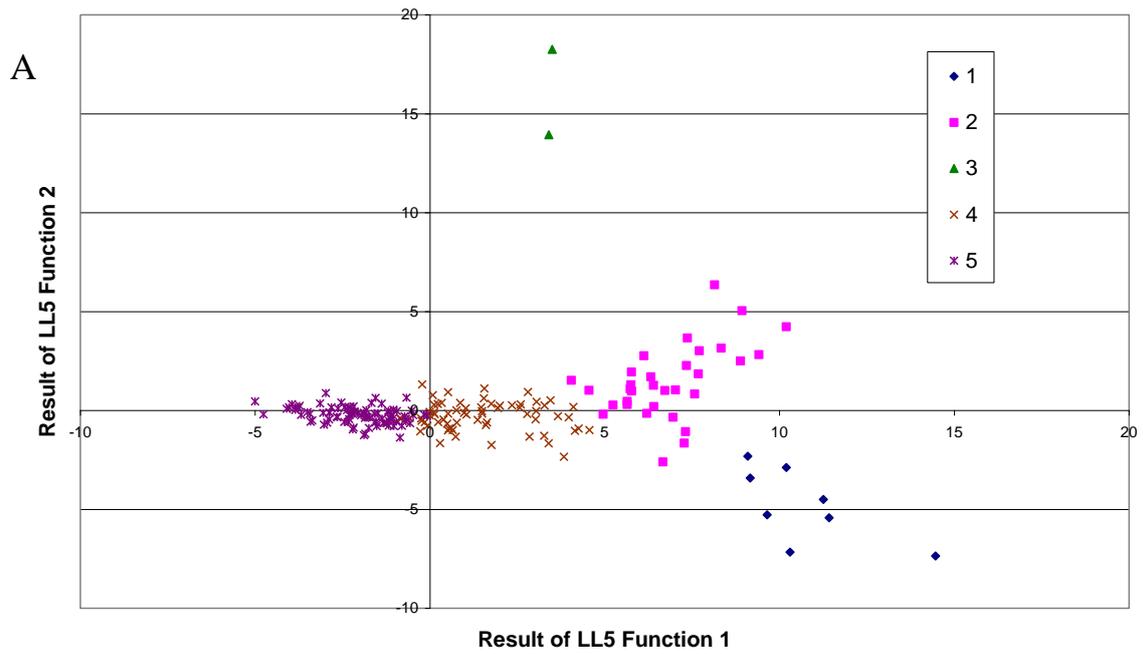


Figure 4.5 RS Clusters with RS data (A) compared to DI data (using RS discriminant functions) (B) for LL5. Clusters are numbered 1 to 5.

4.4 Discussion

The presented data have shown that although measuring colour from digital images is convenient, the colours measured and classified do not correspond very well to those colours measured by reflective spectrophotometry and would have limited use in studies requiring accurate measurement of hair colour, such as genetic association studies. They also do not provide an accurate description of hair colour for crime investigators to utilise as visual identifiers or in description announcements. Some use might be found in distinguishing between a small number of clusters with digital images, as in the two clusters of Analysis 1 (which showed 82.1% of cases correctly classified); however, as the number of clusters or groups increases, there is a lower correspondence. As shown in Chapter 3, observer-reported colours were not well discriminated by RS data (70.4% of cases correctly classified) and were more poorly discriminated by using digital image analysis (52.0% of cases correctly classified). These results support the assertion that objectively defined and measured colours, rather than observer or self-reported colours are necessary to classify individuals for further study and at this time, this is best achieved through reflective spectrophotometric measurement.

It should be noted that the images in this study were all taken with the same camera under the specific conditions of a well-lit room and from a close, specified and identical distance and all treated in the same way. Any additional variability in conditions would contribute to less similarity between the digital image analysis and a reflective spectrophotometry standard. If variability could be reduced, through improvement in photographic conditions or through more accurate correction for varying conditions, results may improve. The act of photographing an object introduces error in colour reproduction due to the nature of the camera and image production and is an approximation of the visual field (whether digital or with film)[62]. Different models of camera may have different abilities to reproduce colour accurately and may improve results. The process of measuring RGB values and converting RGB measurements to CIE L*a*b* colour space may also result in the loss of some colour information.

It may also be that there are components of hair colour that are omitted when measuring colour directly from a digital photograph. This may also part of reason for the higher correlations found by Herbin *et al.*[63] and Takiwaki *et al.*[67] in measuring skin colour by digital image analysis. For example, feature such as hair texture, thickness or lustre

may have a different effect on the analysis of colour from a digital image than they have on reflective spectrophotometry or on human perception. These features would have a lesser, or a different, effect than when measuring skin colour. With these things in mind, the improvement of digital cameras and further research into the digital photography of hair would assist in developing methods that provide a more accurate measurement and classification of hair colour that would be of use to researchers, investigators and forensic scientists.

Chapter 5 will introduce microscopic measurements of hair colour and morphology.

Chapter 5 Microscopic Examination of Hair Colour and Colour Morphology⁴

5.1 Introduction

5.1.1 Microscopic Hair Structure

Hair is commonly encountered as physical evidence at crime scenes [21]. Traditional forensic hair examination involves examining the specimen under a microscope and observing characteristics such as colour, length and diameter of the hair, as well as the distribution of pigment granules and other morphological features of the hair that may assist in determining the source of the hair [21-24].

Hair is an appendage of the skin [21, 73] and consists of a hair *root* and hair *shaft* that grow from a hair *follicle* (Please see Figure 5.1). For a traditional forensic hair examination, as well as for the measurement of hair colour, the hair shaft is the most important part and is covered in more detail in Figure 5.2.

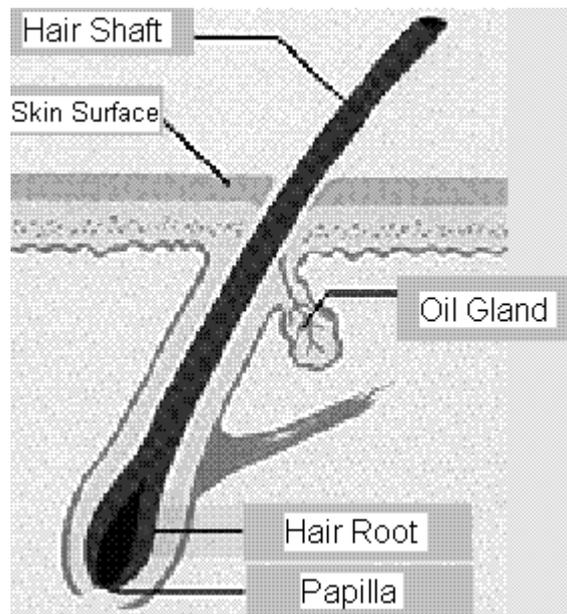


Figure 5.1 A cross section of a hair follicle.

⁴ Parts of this chapter have been published in *Microscopy and Microanalysis*. Vaughn *et al.* 2009:15 Pages 189-193. Please see Appendix 4

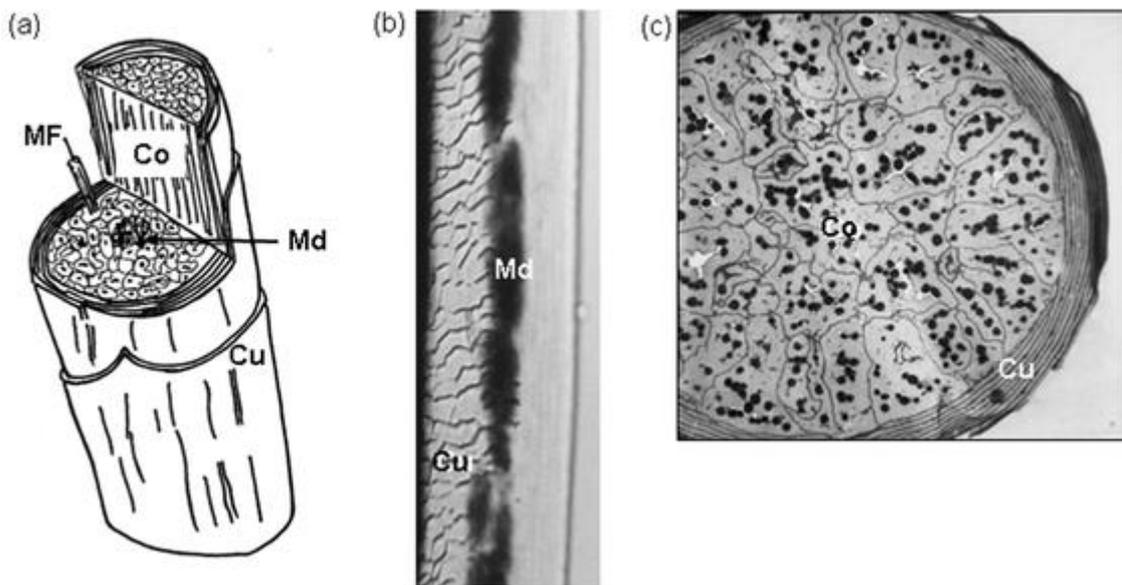


Figure 5.2 Illustrates the microscopic features of the hair shaft. a) Cartoon of the hair shaft with cut-through to reveal inner aspects. Co, cortex; Cu, cuticle; Md, medulla. A macrofibril (300 nm) of single cortical cell is revealed with a constituent microfibril (MF; 7–10 nm). b) Whole mount of a scalp hair shaft with focal plane on cuticle (Cu, left) and medulla (Md; middle). c) Transverse section of a fine hair shaft revealing a multilayered cuticle (Cu) encompassing a tightly packed cortex (Co) containing many dark melanin granules. From Tobin *et al.* 2006 [74].

5.1.2 Cellular Biology of Pigmentation

Melanocytes are dendritic, neural-crest derived cells found in the epidermis, hair follicle and iris [4]. Melanocytes in the hair follicle, specifically melanocytes in the proximal hair bulb [6], produce melanin. Melanin is produced in cellular organelles called melanosomes; these pigment granules are then secreted and taken up by the keratinocytes that make up the growing, pigmented hair shaft [4, 6]. Melanosomes are cellular organelles (similar to lysosomes) that package the melanin and have a role in its distribution [4, 6].

Once melanin is formed in the hair bulb it is passed to surrounding keratinocytes that will be incorporated into the hair shaft [4, 73]. Unlike skin, which is continuously pigmented, pigmentation of hair only occurs during the hair's active growth stage (called the *anagen* stage) [6]. The mechanism for this transfer is uncertain. Several hypotheses have been proposed [6, 75]; the most likely method is by exocytosis of the melanin granules and uptake by the keratinocytes, the other possible mechanisms include cytophagocytosis of the melanocytic dendrite, fusion of the plasma membrane

between the melanocyte and keratinocyte or membrane vesicles that are released and are ingested by the keratinocytes.

Eumelanin (the black/brown pigment) is produced in ellipsoidal melanosomes with a fibrillar matrix, while pheomelanin (the red/yellow pigment) is produced in smaller, spherical melanosomes with a vesiculoglobular matrix and irregular melanin deposition [4, 6, 51]. The switch between eumelanin and pheomelanin production is controlled by the melanocortin-1 receptor (MC1R) and proopiomelanocortin (POMC) ligand pathway [4, 14].

The shape of the melanosomes and the packaging of melanin granules may also influence colour of skin, eyes and hair by scattering light in different ways [4, 73]. In addition, darker pigmentation of skin, and possibly of hair, is associated with higher numbers of melanosomes, although the number of melanocytes remains constant [4, 51]. Melanocytes in black hair follicles contain the greatest number of melanosomes (which are eumelanosomes), while the melanosomes in brown hair bulbs are smaller and those in blonde hair are very poorly melanised [6]. While melanin in the epidermal keratinocytes of skin is almost completely degraded over time, this does not occur in hair; melanin remains “stored” in the hair shaft [6]. This is thought to be part of the mechanism that allows for situations such as dark hair and fair skin.

5.1.3 Microscopic Hair Colour Measurement

In an attempt to assist with the classification of hair colours in a forensic examination, Bednarek [24] illustrated one way of measuring hair colour at the microscopic level; calculating the Red-Green-Blue (RGB) values of a digital image of the hair [24]. The average RGB values for ‘Blonde’ and ‘Brown’ hair were determined.

Recent research has also been conducted on the measurement of hair colour at the microscopic level for determining the contributor of a hair for identification purposes [27]. By using a discriminant analysis to assess colour measurements, Brooks *et al.* [27] were able to correctly assign a hair to a particular individual between 58% and 68% of the time (depending on what colour model was used), based on known microscopic colour measurements of the individuals’ hair [27].

5.1.4 Aims

The aim of this experiment was to determine the association between colours measured at the microscopic level with colour measured at the macroscopic level and to evaluate two morphological properties of hair – thickness and pigment pattern.

Since the macroscopic hair colour is made up of hundreds or thousands of strands of slightly different colours, it is important to know if the microscopic colour of a strand can indicate what the head of hair might look like and vice versa. This would be useful if a strand of hair is found as evidence or for the prediction of both macroscopic and microscopic colour from DNA. Examining morphological properties of the hair, such as thickness [53] and pigmentation patterns, and their associations with colour provides more information on the physical basis of hair for the description of phenotypes.

5.2 Materials and Methods

5.2.1 Subjects and Sampling

From the 202 subjects described in Section 3.2.1, 36 were also included in the microscopic analysis, a number restricted by available samples and time available to use the microscopic equipment. This included two individuals with white hair, three individuals of non-European ancestry and 31 individuals of European ancestry between the ages of 18-35, all with their natural hair colour. Photographs of these individuals' hair can be seen in Appendix 6.

In addition to samples and measurements described in Sections 3.2.2 and 4.2.2, several full-length strands of hair were cut from the scalp of these volunteers.

5.2.2 Preparation of Hairs for Microscopic Analysis

Hairs were mounted on Livingstone Premium glass slides (76.2 x 25.4 mm pathology grade) with Entellan mounting medium (Merck) and covered with Marienfeld cover slips (22 x 60 mm No. 1). This medium has a refractive index of 1.5, which is required for proper viewing under the conditions described in Section 5.2.3. The area to be examined was marked on the slide. Three areas on each of three hairs were examined for each individual and the mean of these measurements was used for most analyses. A hair prepared for microscopic analysis can be seen in Figure 5.3.

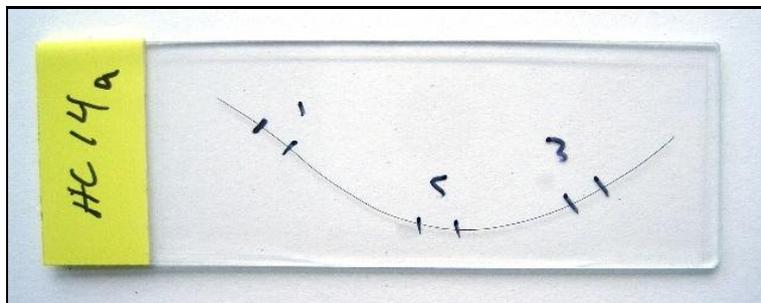


Figure 5.3 Hair prepared for microscopic analysis.

5.2.3 Generation and Collection of Hair Colour Data

Microscope and Image Collection software was generously made available for use at the Australian Federal Police Laboratories in Weston, ACT. Protocols for microscopy and montage were developed and optimised by Elizabeth Brooks for her MSc thesis [27]. Protocols for colour measurement were developed by Bruce Comber (AFP), Elizabeth Brooks (AFP) and myself with some source code for the conversion of RGB to L*a*b* provided by Alasdair McAndrew and Zhenliang Shen (Victoria University, School of Computer Science and Mathematics).

Equipment and Software

Leica DMLB Microscope with a 50x oil immersion lens (500x magnification; field of 276µm), using 10.5V illumination (daylight equivalent) and with DLF (daylight filter) on.

Olympus DP70 12.5 mega pixel digital camera (Olympus Optical Co. Ltd. 2003) and *Image Management System* which consists of the DP Controller (Version 2.1.1.183) and DP Manager (Version 2.1.1.163). An image size of 1360 x 1024 pixels was used with ISO 200 sensitivity, exposure time set to 'automatic', 1% (of the field) contrast correction and all images saved as TIF files . See Figure 5.4

Automontage-Pro 5.02.0096 software (by Syncroscopy; a division of Synoptics Group, UK) was used to montage images of the middle third of the hair (the area containing the greatest amount of pigmentation information) [27]. Hair thickness was estimated by the depth of the optical field. See Figure 5.5

V++ Version 4.0 (by Digital Optics Ltd. New Zealand, ©1990-2002) with modules written to convert RGB colour measurements to L*a*b measurements was used to extract the correct areas of the microscopic images and to measure the colour of the extracted area [64, 69].

Adobe Photoshop CS2 (by Adobe Systems Inc., San Jose, California, USA © 2005)

Microsoft Excel 2003 (by Microsoft Corporation © 1985-2003)

The procedure for the collection of the montage images is as follows. On the back of each slide, three points along the hair were marked (Figure 5.3). The hair was found with a 20x lens and the camera adjusted so the hair ran horizontally across the screen, then the magnification was increased using the 50x lens.

The 1% contrast box was dragged to a position so that half of the box was over the hair and half was over the background. The white balance tool was used by clicking on the background.

The depth of the hair was found by noting the number on the fine focus knob; first when the pigmentary features came into focus (or for very light hairs, when the cortical scales disappeared from focus) and next when the pigmentary features ended (or the cortical scales returned to focus). E.g. $40 - 55 = 15\mu\text{m}$ depth.

The depth of the field was divided by three. This gives the number of optical planes required for the middle third of the hair. From the previous example; $15 \div 3 = 5$ optical planes that start at 45.

Images were captured for the middle third of the hair (e.g. at 45, 46, 47, 48 and 49) and saved to file. These are the 'source images'. The source images were opened in *Automontage-Pro* (Figure 5.6) using the following settings: 'align' source images', 'scan montage', and 'save the montage to file'.

The montaged image was opened in V++ and was 'extracted' by highlighting the hair and removing the image background. Edge artefacts (resulting from light refraction) were removed from the image. If hairs contained a black medulla (an artefact of the slide mounting medium), this was also removed with *Adobe Photoshop* by cutting the top part of the image and moving it over the medulla (Figure 5.7). The new 'extracted montage' image was saved to a separate folder (Figure 5.8 and Figure 5.9)

In V++, the folder with all extracted montages was analysed using the *rgb2lab* function. Since all images were taken under identical lighting conditions, no corrections were made to the resulting L^*a^*b values. The results were copied into an Excel spreadsheet for analysis.

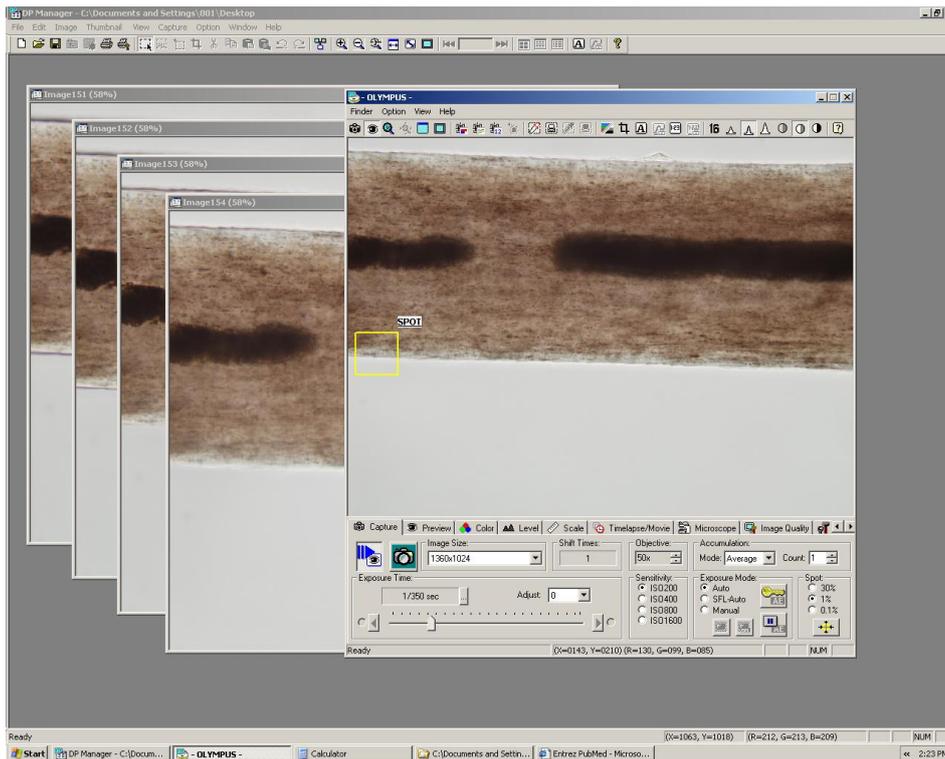


Figure 5.4 Screenshot of the Olympus DP Manager used to take images

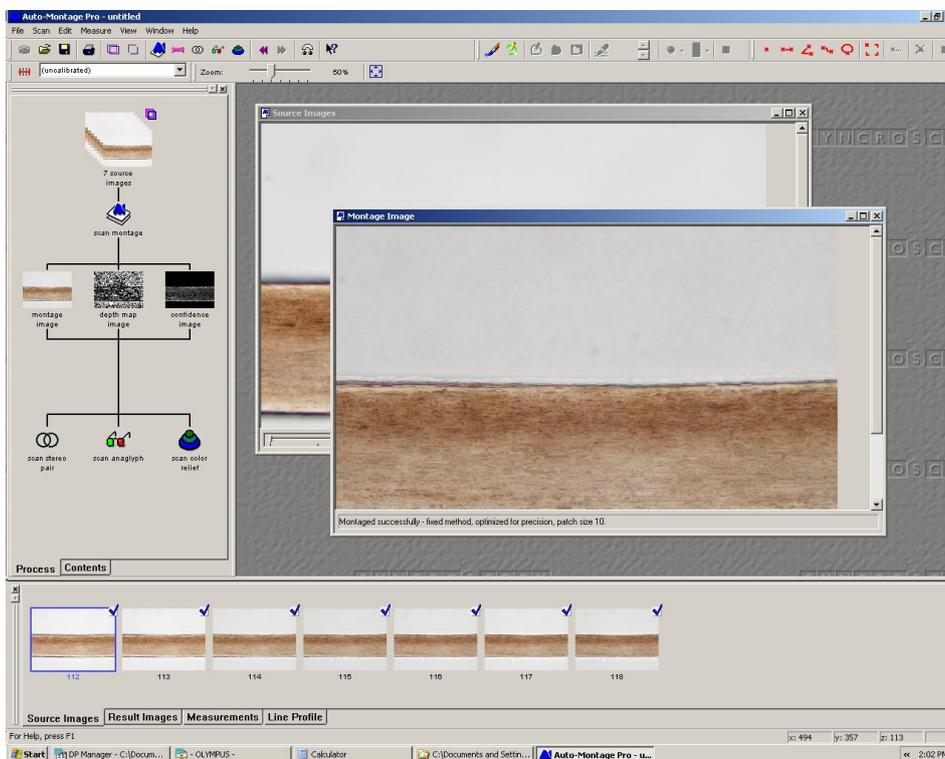


Figure 5.5 Screenshot of the Automontage software

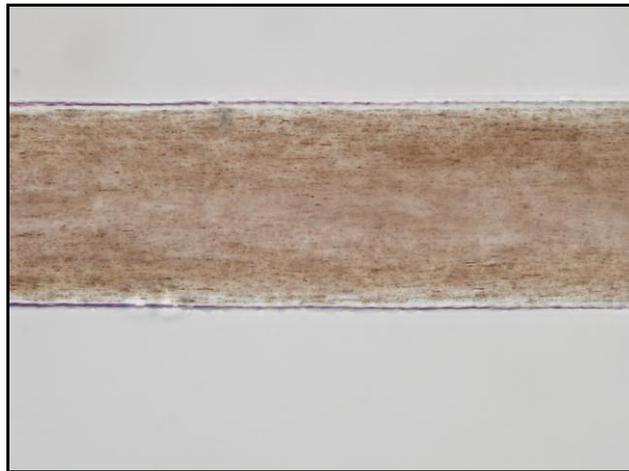


Figure 5.6 – Montaged image

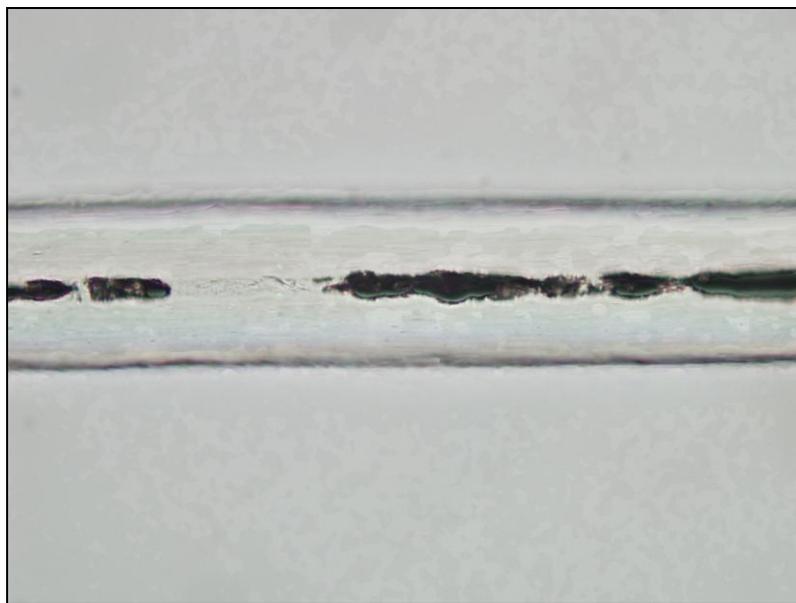


Figure 5.7 Image of a black medulla that was later removed with Adobe Photoshop



Figure 5.8 – Extracted image of a 'Light Brown' hair



Figure 5.9 – Extracted image of a 'Blonde' hair

5.2.4 Generation and Collection of Hair Thickness Data

Hair thickness was estimated by the depth of the optical field, as described in Section 5.2.3. The thickness of the hair may have an effect on the colour by affecting how light travels through the hair and reflects from the pigment.

5.2.5 Pigmentation Pattern

The montage image produced in Section 5.2.3 was transformed into a monochrome image and the standard deviation of image intensity was measured. This measurement gives an estimate of how variable the density of the pigment granules is, which may also have an effect on the colour by affecting how light travels through the hair and reflects from the pigment.

5.3 Results

5.3.1 A Comparison of Microscopic Colour to Reflective Spectrophotometry

When comparing microscopic (Mc) colour L^* , a^* and b^* measurements directly to those determined by RS, regression analysis found that a^* and b^* measurements fail to show any significant linear relationship between RS and Mc (p-values = 0.317 and 0.288, respectively). RS and Mc L^* show a moderate, significant linear relationship with $R = 0.631$ ($R^2 = 0.469$, p-value = 0.000). However; using curve estimation (multiple regression), it can be seen that there is a stronger quadratic relationship between the variables. With RS L^* as the independent variable, $R = 0.751$ ($R^2 = 0.564$, p-value = 0.000). With Mc L^* as the independent variable, $R = 0.904$ ($R^2 = 0.816$, p-value = 0.000). This is a strong relationship and indicates that microscopic images are better at predicting the RS L^* than vice versa. Figure 5.10 illustrates this relationship.

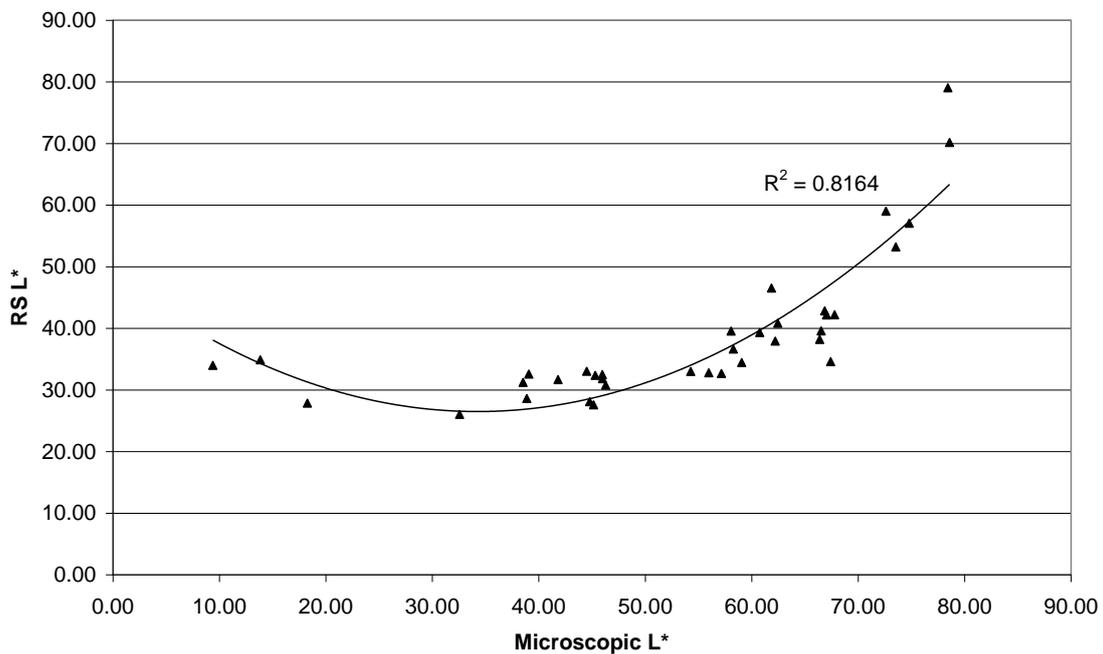


Figure 5.10 The quadratic relationship between microscopic L^* measurements and L^* measurements by reflective spectrophotometry (RS). The trendline has the equation $y = 0.0188x^2 - 1.2854x + 48.505$ and $R^2 = 0.8164$.

5.3.2 A Comparison of Microscopic Colour to Macroscopic Digital Image Analysis

When comparing microscopic (Mc) colour measurements to those determined by digital image analysis of macroscopic hair colour (DI) it was found that a^* and b^* showed no significant correlation between the two methods of measurement, while L^* showed a significant, strong linear relationship between the two, with a correlation coefficient of 0.818 ($R^2 = 0.670$, p -value = 0.000). Multiple regression did not improve the prediction ability of the model.

5.3.3 A Comparison of Microscopic Colours to Reflective Spectrophotometry Clusters

Individuals were assigned to their clusters as determined in Chapter 3; LL2, LL5 and LL6. A discriminant analysis was then performed on these individuals using their microscopic colour measurement data. It was found that the percentage of cases correctly classified decreases as the number of clusters increases with LL2, LL5 and LL6 having 87.9%, 69.7% and 51.5% of cases correctly classified, respectively.

5.3.4 A Comparison of Microscopic Colours to Hair Thickness

Microscopic hair colour was compared to the estimates of hair thickness. There were significant, negative correlations between the microscopic L^* values of hairs and their estimated thickness (in microns) with an $R = -0.705$ (p -value = 0.000). The a^* value was positively correlated to the hair thickness ($R = 0.567$, p -value 0.000) while b^* showed no significant relationship.

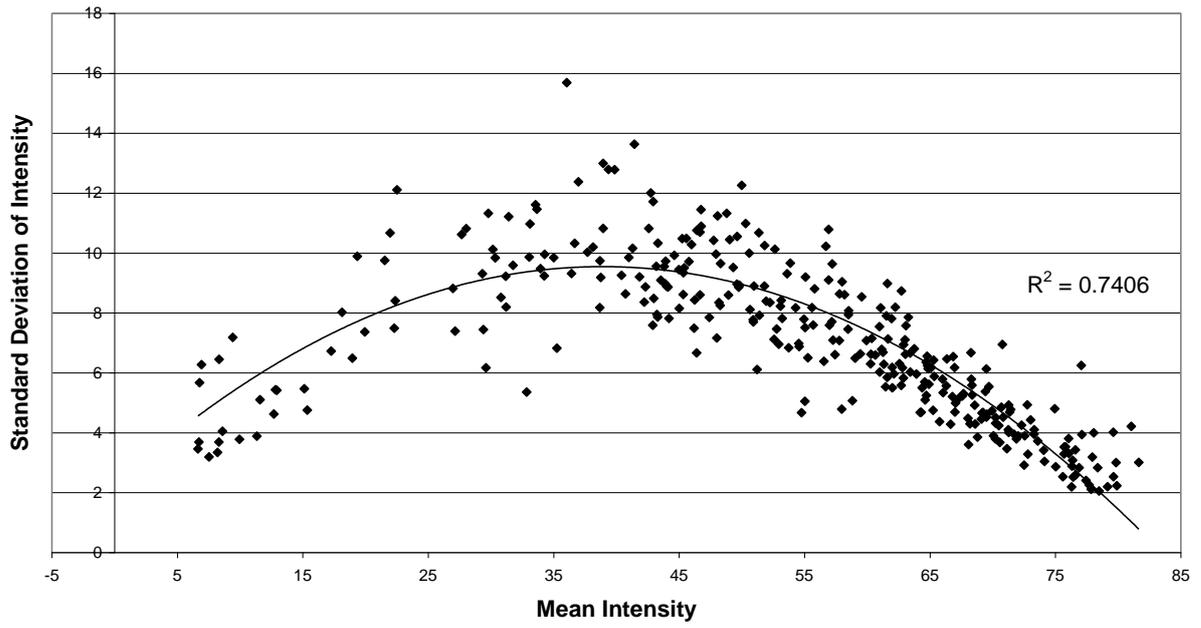
5.3.5 A Comparison of Macroscopic Colour to Hair Thickness

Macroscopic hair colour, measured by reflective spectrophotometry was compared to the estimates of hair thickness. Macroscopic L^* and b^* were found to have a moderate, negative correlation with hair thickness, with correlation coefficients of -0.648 and -0.657 (p -values = 0.000), respectively. The a^* value showed a significant (p -value 0.023) but weaker correlation of -0.390.

5.3.6 Pigment Pattern Analysis

Figure 5.11 shows the results of comparing the standard deviation of the intensities of the monochrome hair images to the mean intensity value by A) image (the mean intensity of all the pixels in an image) and by B) individual (the average intensity of the nine images for each individual). The mean intensity is equivalent to the L^* value for that image. There is an obvious quadratic relationship between the standard deviation and the mean intensity of the monochrome images ($R^2=0.741$, p -value = 0.000) which holds when individuals are examined but is not as strong ($R^2= 0.428$), probably due to a smaller sample size. This indicates that there is less deviation in light hairs that increases as the hair gets darker until a critical point (when intensity is around 40) and then decreases again with darker hairs.

A



B

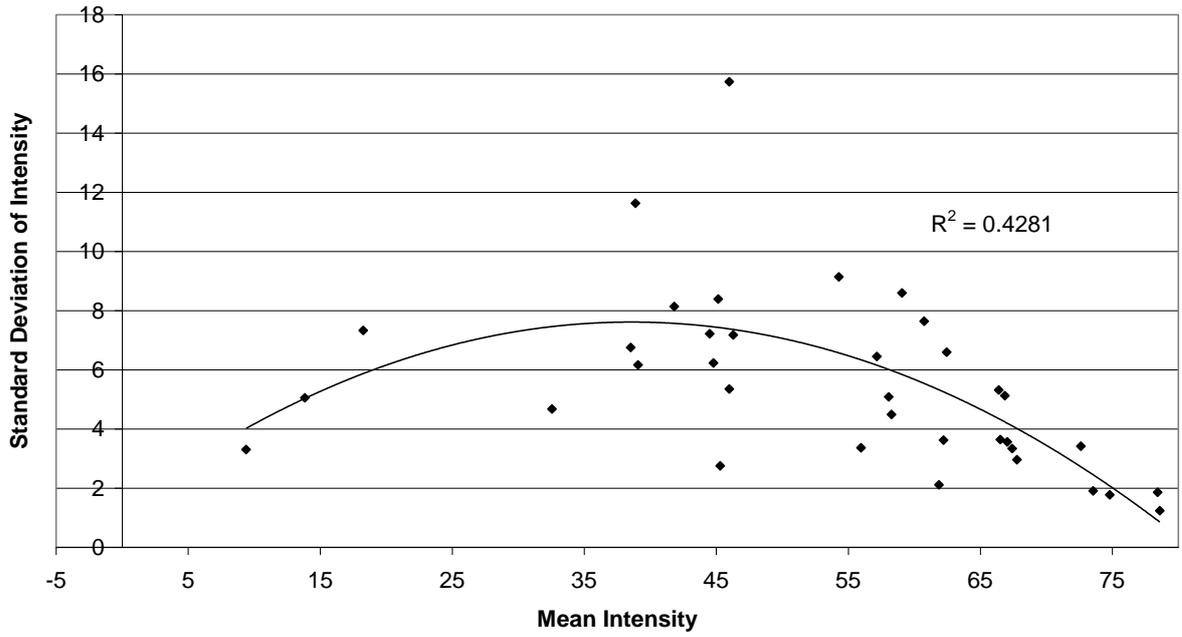


Figure 5.11 The standard deviation of the intensities of the monochrome hair images compared to the mean intensity value by individual (A) and by image (B).

5.4 Discussion

This study illustrates the relationships between macroscopic and microscopic colour, hair thickness and pigmentation pattern. It can be seen that there is a strong relationship between macroscopic and microscopic L^* values that would be useful if one was trying to predict the colour of a head of hair (macroscopic) from a strand (microscopic) or if one was trying to predict hair colour phenotype by genetic analysis and either the macroscopic or microscopic hair colour was known to be more strongly associated with a genetic marker. The relationship between hair lightness (L^*) and thickness is also clearly illustrated at both the macroscopic and microscopic levels.

While the reported relationships are useful and will assist in further research towards the objective measurement of hair colours, it can also be seen that there are weak or insignificant relationships between some aspects of the macroscopic and microscopic colour, particularly when examining the a^* and b^* colour components. These differences may be due to the fact that the number of hair strands that can be examined under the microscope will most likely be a small proportion of the head of hair and large sample sizes are difficult to obtain.

Colour measurement by digital image analysis also presents some problems. In Chapter 4, it was shown that macroscopic hair colour measured by digital images analysis does not correspond very well to the reflective spectrophotometry readings of the macroscopic hair colour and some of the same factors may be involved in measuring the colour of microscopic images; such as the ability of the camera to reproduce colour and the conversion from RGB to the $L^*a^*b^*$ colour space. However; the high degree of consistency in conditions under a microscope should make microscopic digital images of hairs comparable to each other and therefore more consistent in their relationships with the macroscopic colour.

The measurement of hair thickness could also be improved to increase the strength of relationships between thickness and colour. Cross-sections of the hair may be examined to determine the shape before measuring the thickness of hairs, since using the optical

depth may inadvertently use different sides of an oval-shaped hair for measurement and therefore be less precise.

The microscopic examination of the pigmentation patterns shows that hair has a smooth pigmentation distribution that becomes more variable as the hair gets darker until an intensity of about 40 and then gets smoother again (because of the saturation of pigment in the hair shaft). These results may indicate that the structure of the hair's matrix and the distribution of pigment granules in the hair shaft (which would also be affected by the hair's thickness) have an effect on the colour of the strands and this type of information will be important when considering which areas of the genome to examine when looking for genes that relate to hair pigmentation. For example, when studying hair colour, it may be worth looking at genes that control hair thickness since colour and thickness are correlated. Measuring additional pigmentation pattern variables, such as those considered by forensic hair examiners (Figure 2.1) may provide more information about the relationship between pigment distribution and colour.

Chapter 6 will introduce a chemical characterisation of the hair samples.

Chapter 6 Chemical Analysis of Hair Pigments

6.1 Introduction

6.1.1 Chemistry

Melanin is the biopolymer pigment that is the most directly responsible for human pigmentation [4]. While haemoglobin and dietary carotenoids contribute somewhat to the appearance of epidermal colour, they are minor constituents and have little to do with the pigmentation of hair [4].

Melanin is a polymer of quinone and indole-quinone derived units produced from the amino acid tyrosine [4, 20]. Its exact composition is unknown and the information available is inconsistent [76]. Briefly, melanin comes in several forms; two of which are relevant to human pigmentation; eumelanin and pheomelanin. Eumelanin is a black/brown pigment, while pheomelanin is a red/yellow pigment [4, 6, 51]. Studies have shown that it is both the *amount* of melanin present and the *type* of melanin that gives hair its colour [4]. Figure 6.1 shows the structure of the eumelanin and pheomelanin polymers while Figure 6.2 indicates the biosynthetic pathways of each, showing the subunits and the main enzymes involved in the synthesis of melanin, called melanogenesis.

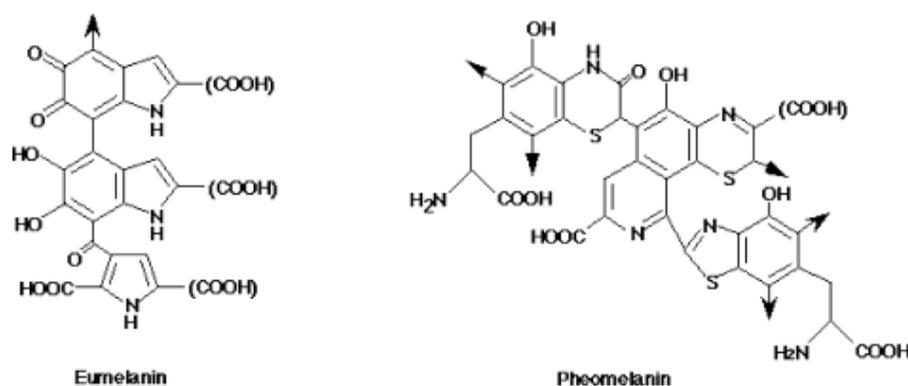


Figure 6.1 The structures of the eumelanin and pheomelanin polymers [76].

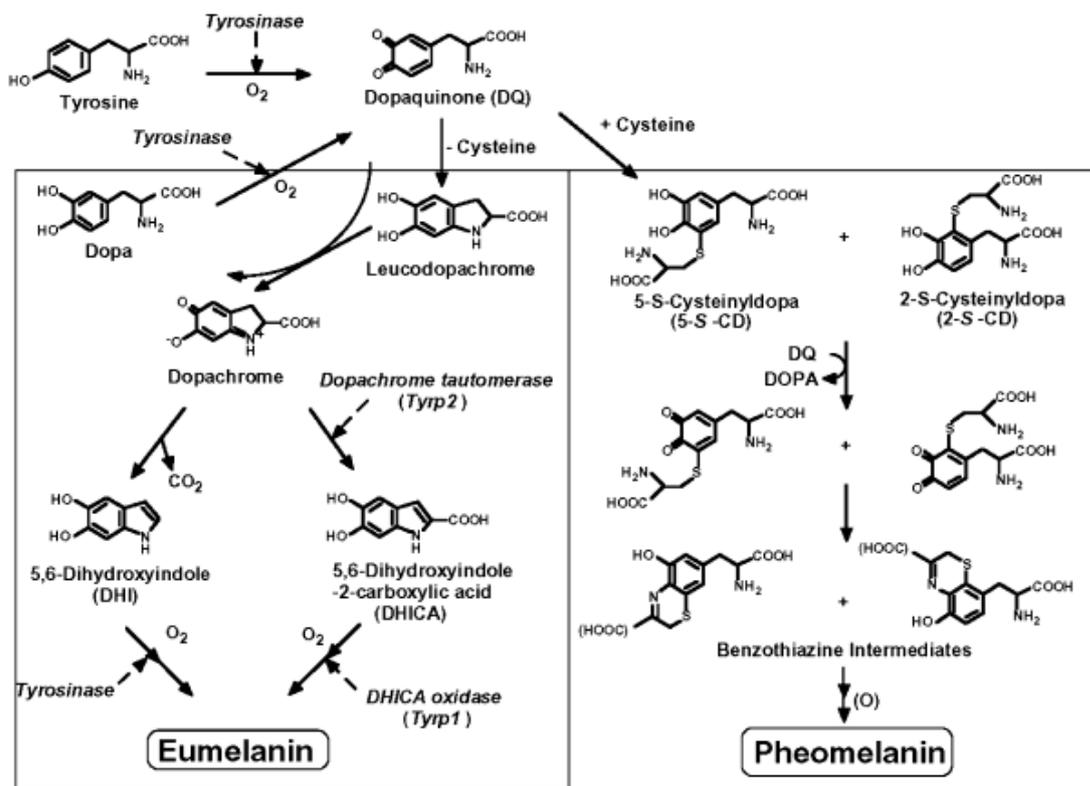


Figure 6.2 The biosynthetic pathways to eumelanin and pheomelanin as established in human skin and in murine hair pigmentation [76].

As mentioned, the exact composition of the polymers is not well understood and not necessarily stable. Eumelanin is composed of 5,6-dihydroxyindole (DHI) and 5,6-dihydroxyindole-2-carboxylic acid (DHICA) derived units [76, 77]. Some hair samples [20] have been characterised as having a eumelanin composition of approximately 60% DHI-derived and 40% DHICA-derived units, however, caution is suggested in applying this result to all hair samples, given the problems in analysis of melanin and the biosynthetic mechanisms of melanin production that was discussed in Section 5.1.2. With the same caution, pheomelanin in hair is a polymer of approximately 80% 2-S-Cysteinyl-dopa (2-S-CD) derived and 20% 5-S-Cysteinyl-dopa (5-S-CD) derived units [20]. The cysteine residues (containing sulphur) give pheomelanin its red or yellow colour. In some animal breeds (e.g. Red Setters) pheomelanin appears red, while in others (e.g. Yellow Labrador Retrievers, certain mouse strains), pheomelanin appears yellow (See Box 6.1) [4].

Black hair has been shown to contain approximately 99% eumelanin and 1% pheomelanin, while brown and blonde hair contain approximately 95% eumelanin and

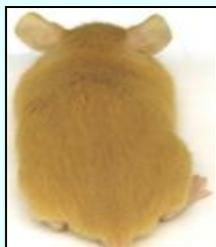
5% pheomelanin, and red hair contains 67% eumelanin and 33% pheomelanin, this unique ratio giving it its distinct, red hue [20].

The exact melanin content of blonde hair has not been examined fully and is still contentious. However the 2001 study by Borges *et al.* [20] indicates that blonde is a result of less overall melanin, rather than a result of either an increased proportion of or a yellow version of pheomelanin (See Box 6.1), while red hair is a *change* in pigment composition. A 2005 study on the photo lightening effects of visible and ultraviolet light by Takahashi *et al.* [17] provides additional evidence for this. In this study, blonde hair was found to be much more sensitive to visible wavelengths of light, while red hair, known to contain a high amount of pheomelanin, was found to be more sensitive to UV light. The differential sensitivity of red and blonde to different wavelength ranges could indicate a different pigment composition.

While melanin is the pigment responsible for hair colour, it must be first synthesised by the cell and transferred to the hair. This cellular biology process was discussed in Section 5.1.2.

Box 6.1

Some animals have a yellow version of pheomelanin. The mice below have approximately the same amount of melanin but the yellow mouse is almost entirely pheomelanogenic while the brown mouse is eumelanogenic [47]. Human 'yellow' or blonde hair is the result of less overall melanin [20, 47].



A^y/a genotype at the agouti locus
yellow



Non-agouti (a/a), $Tyrp1^b/Tyrp1^b$
brown

Photo source: Oetting WS, Montoliu L, Bennett DC. *Color Genes*. (Dec 2008). European Society for Pigment Cell Research. World Wide Web (URL: <http://www.espcr.org/micemut>)

6.1.2 Hair Pigment Analysis; Methods, Uses and Issues

There are several issues to be aware of when analysing or interpreting the pigment content data of hair. Melanin is an extremely difficult compound to analyse, primarily because it is a non-homogeneous biopolymer that is difficult to isolate [76, 77]. Therefore; degradation products are analysed instead, which presents new problems. The experiments by Ito and Fujita (1985) quantified pyrrole-2,3,5-tricarboxylic acid (PTCA) as the degradation product of eumelanin after a permanganate oxidation using HPLC-UV and quantified 3-amino-4-hydroxyphenylalanine (3-AHP) and 4-amino-3-hydroxyphenylalanine (4-AHP) as the degradation product of pheomelanin by HPLC-ECD (both isomers of AHP were detected, but only 3-AHP was quantified because a standard was commercially available)[77]. Factors were calculated to convert the amount of the degradation product to the amount of melanin. These conversion factors were based on the amount of PTCA and 3-AHP recovered from the degradation of sepia (cuttlefish) melanosomes (2% and 20% respectively) which gave conversion factors of 50x for PTCA and 5x for 3-AHP [77].

Later experiments in eumelanin analysis used an alkaline hydrogen peroxide degradation method that resulted in significant amounts of pyrrole-2,3-dicarboxylic acid (PDCA) being recovered and quantified, as well as PTCA [78]. These studies showed that the degradation of DHICA-derived units resulted in a 14.1% yield of PTCA and no PDCA, while the degradation of DHI-derived units resulted in a 4.82% yield of PDCA and a 0.01% yield of PTCA. Data from permanganate oxidation of eumelanin [76] also shows that PTCA primarily results from the degradation of DHICA-derived units.

Wakamatsu (2004) recommended a conversion factor of 160x for PTCA in hair due to a yield of 0.6% obtained when the concentration of PTCA in solution was measured by spectrophotometry. The low yield found in human hairs (compared to 2.2% in mice and 2.5% in sheep) indicate that most of the eumelanin in human hairs is not DHICA-derived and therefore indicates a lower activity of Tyrosinase-related protein Type 2 (TYRP2 or DCT)[76]. The pathway by which DHICA-derived units are produced by TYRP2 is illustrated in Figure 6.2. This lack of DHICA-derived eumelanin units is supported by a 2004 study by Commo *et al.* which found no TYRP2 protein in the

melanocytes of the human hair bulb and a barely detectable TYRP2 mRNA amplicon signal [50]. It makes sense to find little or no DHICA-derived eumelanin in hair when the enzyme that produces it (TYRP2) is not present in hair bulbs. This is in contrast to the strong presence of the enzyme and its products found in human hair follicle and skin melanocytes.

The difference in hair eumelanin compared to skin, mouse and sepia melanin presents problems for the conversion of values from PTCA yields to amounts of eumelanin and for the comparison of amounts between these tissues and species. Although the exact composition of hair melanin is unknown [79] and each method of degradation (permanganate or alkaline hydrogen peroxide) may have different yields. Until more is known about the various degradation processes, it is better to use raw PTCA values (per milligram of hair) for comparison purposes and to be conservative when comparing results between studies that use different methods.

In 2002 pheomelanin analysis, Wakamatsu *et al.* were able to separate the 3-amino-4-hydroxyphenylalanine isomer (3-AHP; also called 3-aminotyrosine or 3-AT) and the 4-amino-3-hydroxyphenylalanine isomer (4-AHP) and found that the 4-AHP isomer was a more specific marker for the degradation of pheomelanin [80]. Using this more specific marker, a factor of 9x is used to calculate the amount of pheomelanin from the amount of 4-AHP recovered. There are fewer reported methods for pheomelanin degradation than for eumelanin, so studies should be more comparable, however; the use of 4-AHP values, rather than pheomelanin conversions is likely to allow a more reliable comparison between studies.

A significant issue in the quantification of PTCA, PDCA and 4-AHP is the lack of a commercially available analytical standard. Only 3-AHP may be purchased and even this standard has no information about its solubility, pKa or other chemical properties. PTCA, PDCA and 4-AHP must be synthesised in-house or received from another lab. While Professor Wakamatsu (Fujita Health University, Japan) has been very generous with gifts of these standards, this lack of information, limited quantities and inconsistent availability makes optimisation of methods and new investigations difficult.

Developments in the quantification of biological pigments continue. A method for the quantification of PTCA and PDCA in skin by LC/MS/MS has been developed [81] and new degradation products for the presence of pheomelanin are being investigated [82].

6.1.3 Aims

The aim of this experiment was to chemically characterise the samples collected and to quantify any associations found between the amount of eumelanin and pheomelanin in hair and colour measurements taken at the macroscopic and microscopic level.

Knowing the chemical composition of the pigments in hair will assist in accurately determining of hair colour and will help to direct the decision of exactly which genes to examine in the search for associations. For example, two persons whose hair has same amount of melanin could have a different hair colour - indicating that genes outside the melanogenesis pathway may be of interest (such as hair thickness or pigment distribution).

6.2 Materials and Methods

6.2.1 Sampling

From the 202 individuals described in section 3.2.1, 58 were included in this analysis (those with hair samples available). This group includes the 36 individuals described in 5.2.1 (including three non-Europeans and two white-haired individuals) and 22 more individuals from the main sample. Approximately 30mg of hair was trimmed from each subject for chemical analysis. These samples were cut from the ends of the subjects' hair.

6.2.2 Chemicals and Buffers Used in Eumelanin Analysis

1.05 M Sodium Hydroxide (NaOH)	Ajax Finechem (Seven Hills, NSW)
30% Hydrogen Peroxide (H ₂ O ₂)	Ajax Finechem
5% Sodium Sulfite (Na ₂ SO ₃)	Sigma (Perth, WA)
6M Hydrochloric Acid (HCl)	Ajax Finechem
0.01M Phosphate buffer, pH 2.1	Ortho-phosphoric acid from Ajax Finechem
100% Methanol	Merck Chemicals (Kilsyth, VIC)
0.1% phthalic acid	Aldrich Chemical Co. (Milwaukee, WI)

1000 ng/mL PTCA and PDCA standards (a generous gift from Professor Kazumasa Wakamatsu from the Department of Chemistry, Fujita Health University School of Health Sciences, Toyoake, Aichi, Japan)

6.2.3 Hydrogen Peroxide Degradation of Hair for Eumelanin Analysis

Sample preparation and degradation was performed according to the method described in Napolitano *et al.* (2000) with some differences [82]. The hair was ground in a mortar and pestle with liquid nitrogen (instead of mincing with scissors) to make handling easier and reduce loss of sample. Approximately 5mg of hair from each individual was weighed and put into a glass test tube. 950µl of 1.05M sodium hydroxide and 50µl of 30% hydrogen peroxide was added to the sample, resulting in concentrations of 1M NaOH and 1.5% H₂O₂. The test tubes were covered with Para-film and shaken for 24

hours on a Griffin Flask Shaker at room temperature. After 24 hours, 200µl of 5% sodium sulfite (Na₂SO₃) was added to decompose remaining H₂O₂ as in Borges *et al.* (2001) instead of hydrogen sulfite [20]. Samples were acidified to a pH less than 4 with 143µl of 6M hydrochloric acid (HCl), resulting in a total volume of 1.343mL. Samples were then transferred to 1.5mL microcentrifuge tubes and centrifuged at maximum speed for 2 minutes. The supernatant was transferred to HPLC vials for analysis.

6.2.4 HPLC Conditions for Eumelanin (PTCA) Analysis

HPLC analysis was performed using Varian HPLC autosampler (Module 9100), pump (Module 9012) and UV detector (Module 9050) set to 270nm [83]. An Alltech Adsorbosphere 5µm 260 x 4.6 mm column (with matching guard column) was used and heated to 55°C with a Biorad HPLC Column Heater. Mobile phase A was 0.01 M Phosphate Buffer (pH 2.1) and Mobile phase B was 100% methanol. Mobile phases were run in a gradient as described in [20]: A:B at 98:2 ramped evenly to 40:60 over the first 14 minutes, held at 40:60 for 6 minutes, ramped back to 98:2 over the next 5 minutes and held at 98:2 for the last 5 minutes. The total run time was 30 minutes plus a 1 minute equilibration.

6.2.5 Chemicals and Buffers Used in Pheomelanin Analysis

3-amino-4-hydroxyphenylalanine standard (3-aminotyrosine)	Sigma (Perth, WA)
HPLC Grade water with 0.1% formic acid	Riedel-deHaen (Germany)
20mM borate buffer (pH 9.5)	Boric acid from Sigma (Perth, WA)
30mM sodium cyanide (NaCN)	Ajax Finechem (Seven Hills, NSW)
20mM naphthalene-2,3-dicarboxaldehyde (NDA)	Fluka (Milwaukee, WI)
Acetonitrile	BDH Chemicals (Biolab, Clayton, VIC)
50mM ammonium acetate buffer (pH 6.8)	Merck Chemicals (Kilsyth, VIC)

500 ng/mL each (mixed) 3-AHP and 4-AHP standards (a generous gift from Professor Kazumasa Wakamatsu from the Department of Chemistry, Fujita Health University School of Health Sciences, Toyoake, Aichi, Japan)

6.2.6 LC/MS Conditions for Pheomelanin Analysis

Small quantities of AHP are expected from the degradation of pheomelanin in hair and there is no data available on the UV absorption of AHP; therefore, more sensitive alternative detectors are employed in this analysis, such as electrochemical detection (ECD) [77]. Since ECD was not available for this experiment, an attempt was made to quantify 3-AHP with Liquid Chromatography – Mass Spectrometry (LC/MS) according to methods adapted from Nezirevic *et al.*(2007) [84].

The isocratic mobile phase was acetonitrile with 0.1% formic acid (pH 3.0). A ZIC-HLIC® Column (150mm x 2.1mm) was used (SeQuant, Umea, Sweden) with a Finnigan Surveyor Autosampler Plus, LC Pump Plus and LCQ Deca XP MAX module mass spectrometer (Thermo Scientific Fisher, Waltham, MA) using electrospray ionisation.

While the LC/MS method using a ZIC-HLIC® Column worked with a 3-AHP 1ppm standard, the limit of detection of this instrument (approximately 0.1ppm) was too high to be of use in the quantification of AHP from hair samples (data not shown) and was therefore, not used further.

6.2.7 HPLC Conditions for Pheomelanin using Fluorescent Detection

Methods used for the fluorescent detection of 3-AHP and 4-AHP (for the analysis of pheomelanin) were performed according to a previous study by Yang *et al.* (2007) [85].

First, the 3-AHP standard was put through a pre-column derivatisation [85]. 10µL of the 3-AHP standard was added into 1 mL of borate buffer (20 mM, pH9.5). After the addition of 30µL sodium cyanide (30 mM), the solutions were left to equilibrate at room temperature for 10 min. Then 10µL NDA (20 mM) solution in methanol was added and the mixture equilibrated for 30 min at 25 °C.

Chromatography was performed on the same Varian HPLC setup as described in Section 6.2.4, but with a Varian 9070 Fluorescence detector with a PMT voltage of 700.

Mobile Phase A (95%) was acetonitrile and Mobile Phase B (5%) was 50mM ammonium acetate (pH 6.8). A Supelco APS-2 Hypersil aminopropyl column was used (250mm x 4.6mm, 5 μ m particles) with a matching guard column. Fluorescence was detected with an excitation wavelength of 442nm and an emission wavelength of 482nm as determined by the Perkin Elmer Fluorescence Spectrophotometer LS50B.

This HPLC-FD method worked well with the 3-AHP standard and had a very low limit of detection; however the method was unable to resolve the 3-AHP and 4-AHP from a mixed standard or from actual samples. This may be due to a degradation of the standards over time (standards are not available commercially and were a gift from Professor Kazumasa Wakamatsu from the Department of Chemistry, Fujita Health University School of Health Sciences, Toyoake, Aichi, Japan) or due to the fluorescence of contaminants and undesired compounds. There is no published data on the stability of standards; therefore fresh standards and additional optimisation of this method may provide results in the future.

6.3 Results

6.3.1 Method Validation for Eumelanin (PTCA) Analysis

Standards for PTCA and PDCA (1000ng/mL) were a generous gift from Kazumasa Wakamatsu from the Department of Chemistry, Fujita Health University School of Health Sciences, Toyoake, Aichi, Japan. 1000ng/mL is equal to 1000 parts per billion (ppb) or 1 part per million (ppm). Phthalic acid was used during method optimisation since it has a retention time very near the desired compounds (Borges *et al.*, 2001), however it was not used to quantify samples as it was found to persist through subsequent injections.

Standard curves were prepared from 50ng/mL to 1000ng/mL for PTCA and from 100ng/mL to 1000ng/mL PDCA; 50ppb and 100ppb being the lower limits of detection for PTCA and PDCA respectively. Each concentration was repeated twice and R^2 values for each standard curve were excellent; R^2 for PTCA was 0.9995 and for PDCA was 0.9951 (please see Figure 6.3 and Figure 6.4).

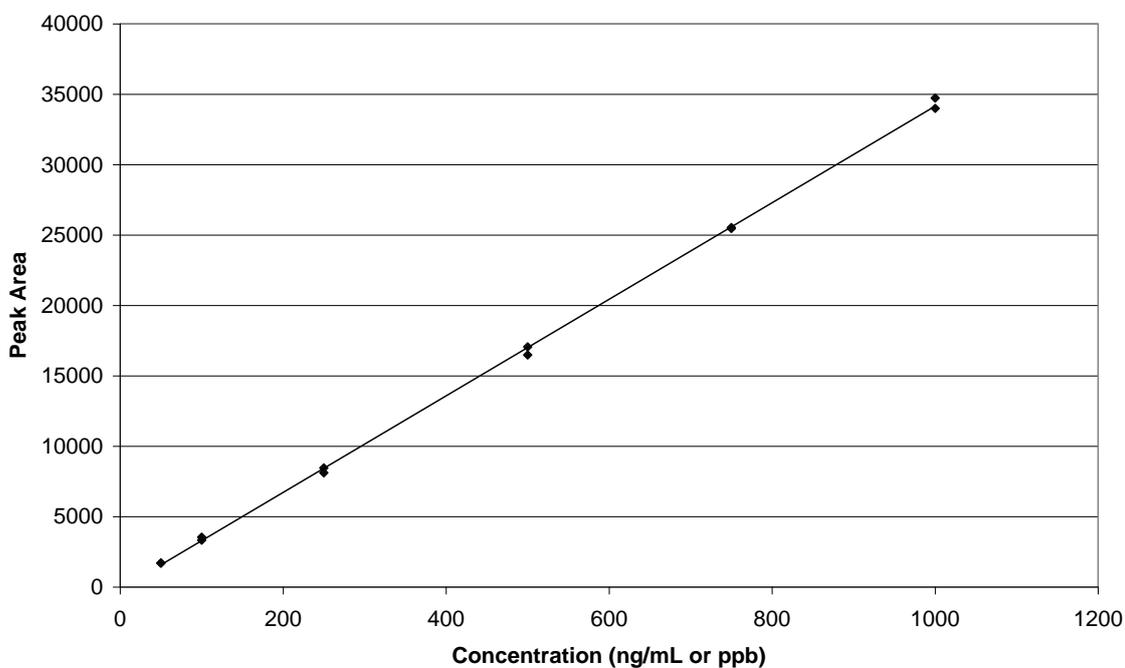


Figure 6.3 Standard curve for PTCA ($R^2 = 0.9995$) where $y = 34.301x - 138.04$

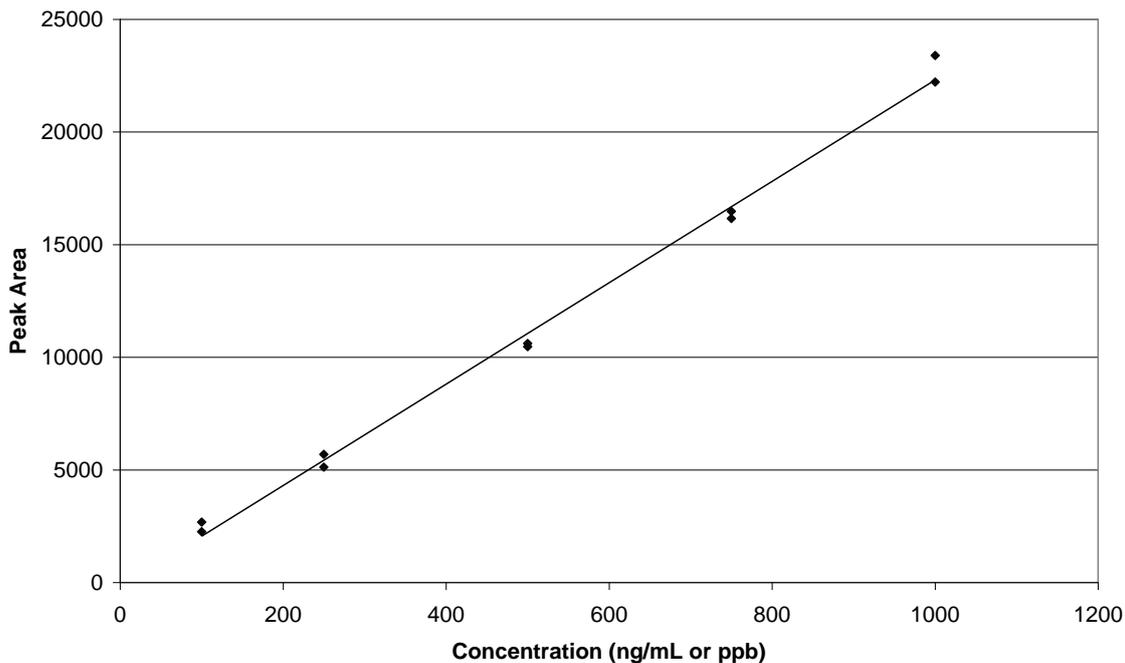


Figure 6.4 Standard curve for PDCA ($R^2 = 0.9951$) where $y = 22.499x - 192.71$

Retention times were approximately 9.4 minutes for PTCA and 8.3 minutes for PDCA. Degraded samples of dark and light hair were performed and analysed and typical chromatograms are shown in Figures 6.5 and 6.6.

Although PDCA was detected in the dark sample (black hair), the concentrations were so low that it was not worthwhile to continue quantifying this compound. It was not found in the fair hair at all. Blank samples showed no detectable PTCA or PDCA. Chromatograms were automatically integrated from 7 to 14 minutes.

To calculate the amount of PTCA for each milligram of hair, the following formula was used from the standard curve:

$$\text{Concentration (ng/}\mu\text{l)} = (\text{Peak Area} + 138.04)/34.301$$

$$\text{Concentration (ng/}\mu\text{l)} \times 1.343\text{mL} = \text{Amount of PTCA (ng)}$$

$$\text{Amount of PTCA (ng)}/\text{Amount of hair (mg)} = \text{ng PTCA}/1 \text{ mg hair}$$

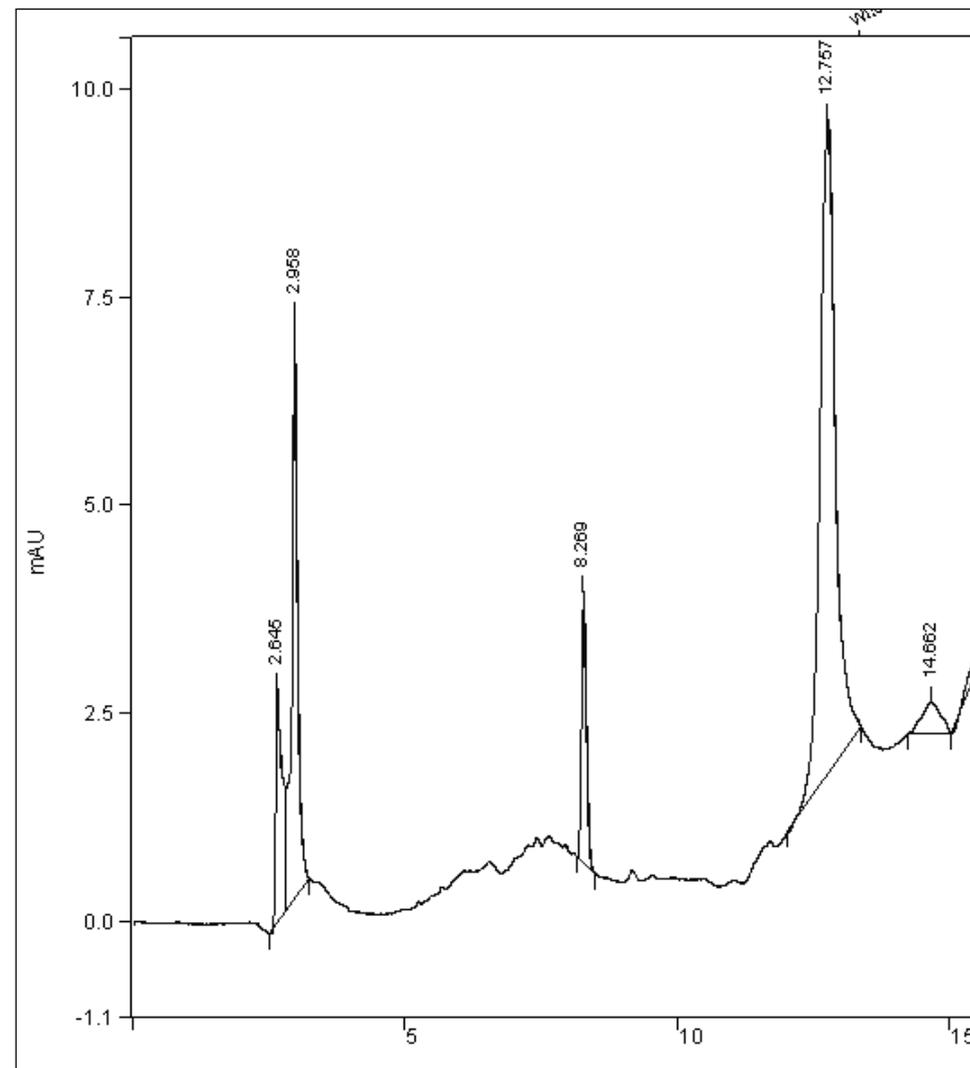
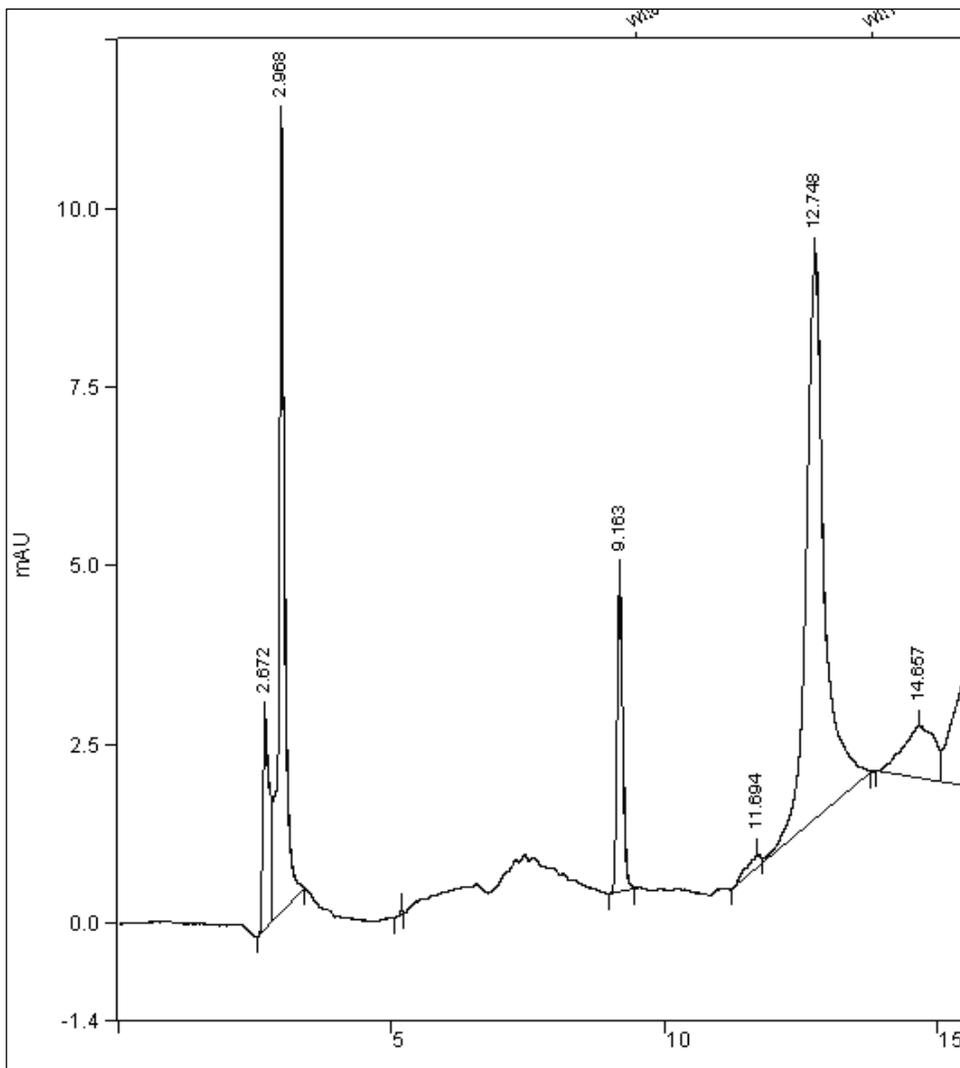


Figure 6.5 Sample Chromatograms

A – 500ppb PTCA Standard (9.391) and showing phthalic acid (12.748)

B – 500ppb PDCA Standard (8.269) and showing phthalic acid (12.757)

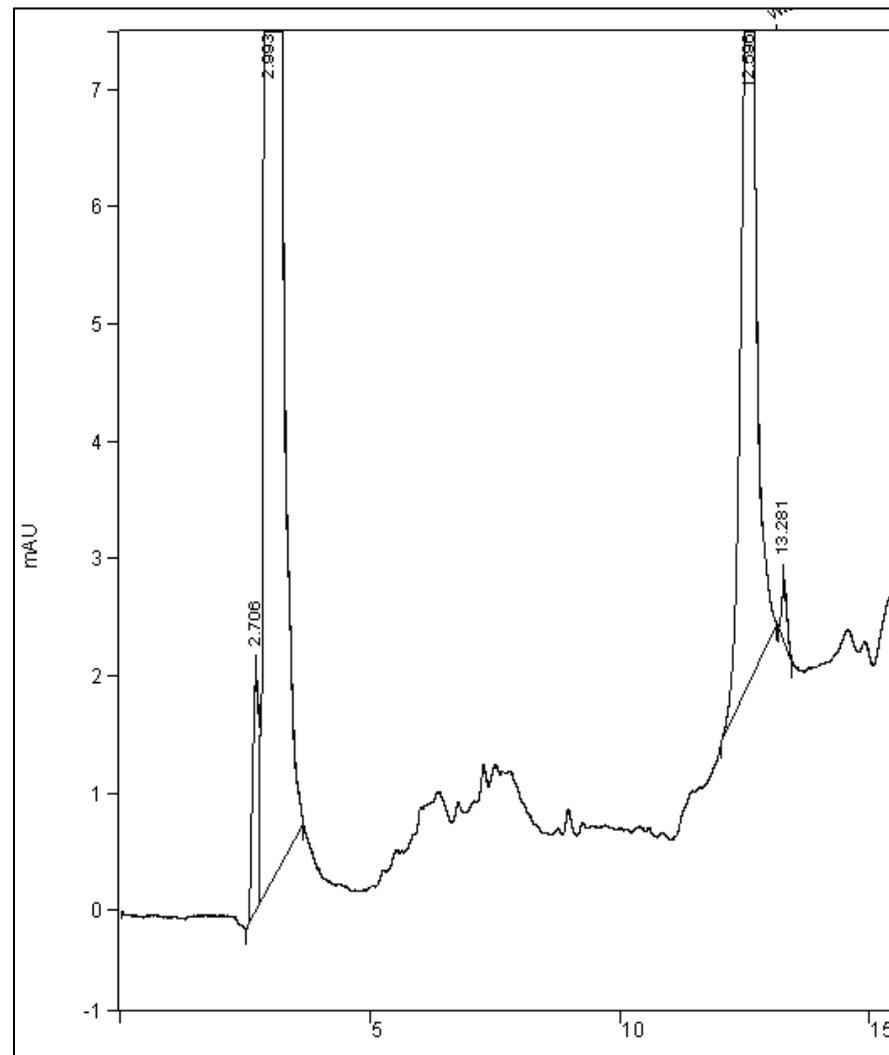
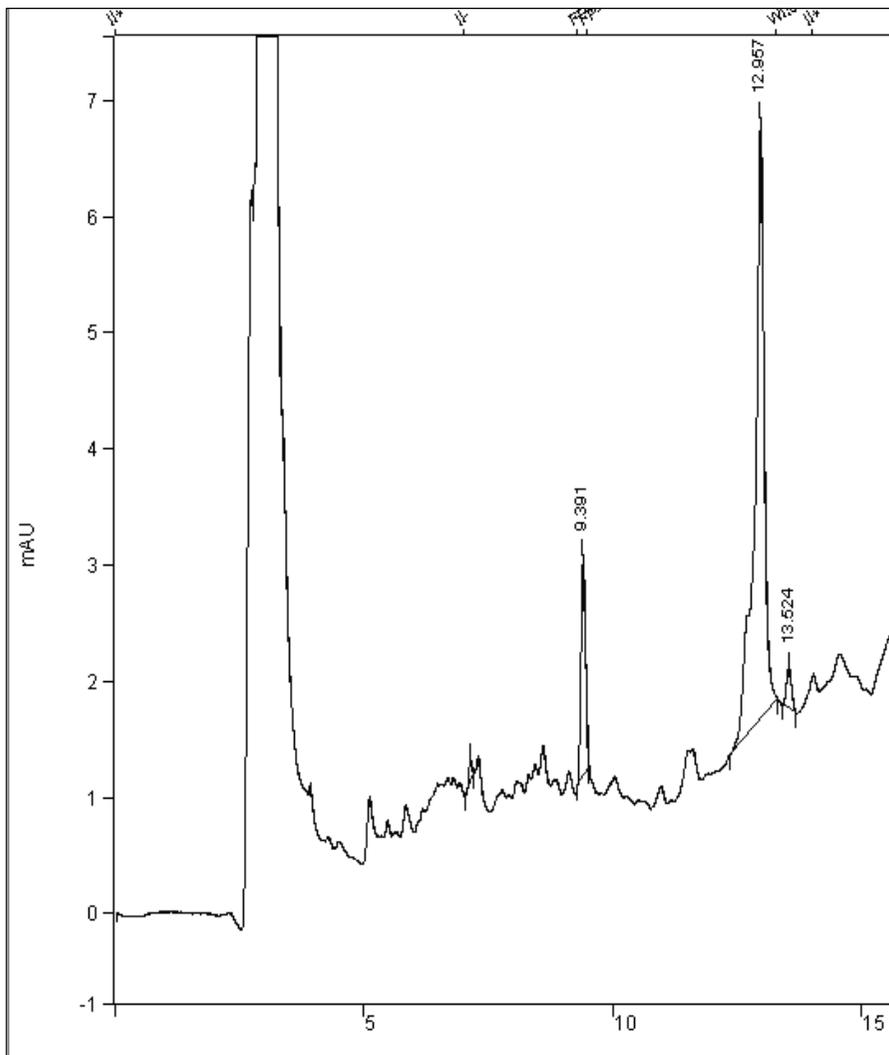


Figure 6.6 Sample Chromatograms

C – A typical chromatogram for dark hair showing PTCA (9.391) and phthalic acid carry-over.

D – Blank sample (no hair in degradation) showing no PTCA but with phthalic acid.

For each run, a 500ppb standard was also analysed and if the calculated amount of PTCA from the standard curve was within 20% of the known amount, the run was considered valid. All runs showed a less than 4% variation for the 500ppb standard.

To validate the method with respect to variability and linearity, the following tests were performed. Tests of intra-assay variability and inter-assay variability were carried out on light and dark hair and evaluated according to the percent coefficient of variance (%CV) which is the standard deviation as a percentage of the mean. A mean of the results for two injections was used for all calculations. These results can be seen in Table 6.1.

Results for multiple injections were excellent (10% is to be expected) and overall variability for the method at 12% is acceptable and in line with the published results [20, 81]. Variation was consistent no matter if the hair was dark or light.

Table 6.1 Variability in the hair degradation and HPLC method

Sample	Intra-Assay Variability (%CV)	Inter-Assay Variability (%CV)
	8 injections of one degradation	6 Degradations of one hair sample
Dark Hair	6.3%	12%
Fair Hair	6.5%	12%

For PTCA, a test of linearity of recovery was conducted. Various amounts of dark hair were degraded to ensure that there is a linear relationship between the amount of hair degraded and the amount of PTCA recovered. A Pearson's test for correlation showed that the amount of hair and amount of PTCA were found to be significantly correlated ($p=0.02$). The variation in the final results was 9.3%CV for this test. The results are displayed in Figure 6.7.

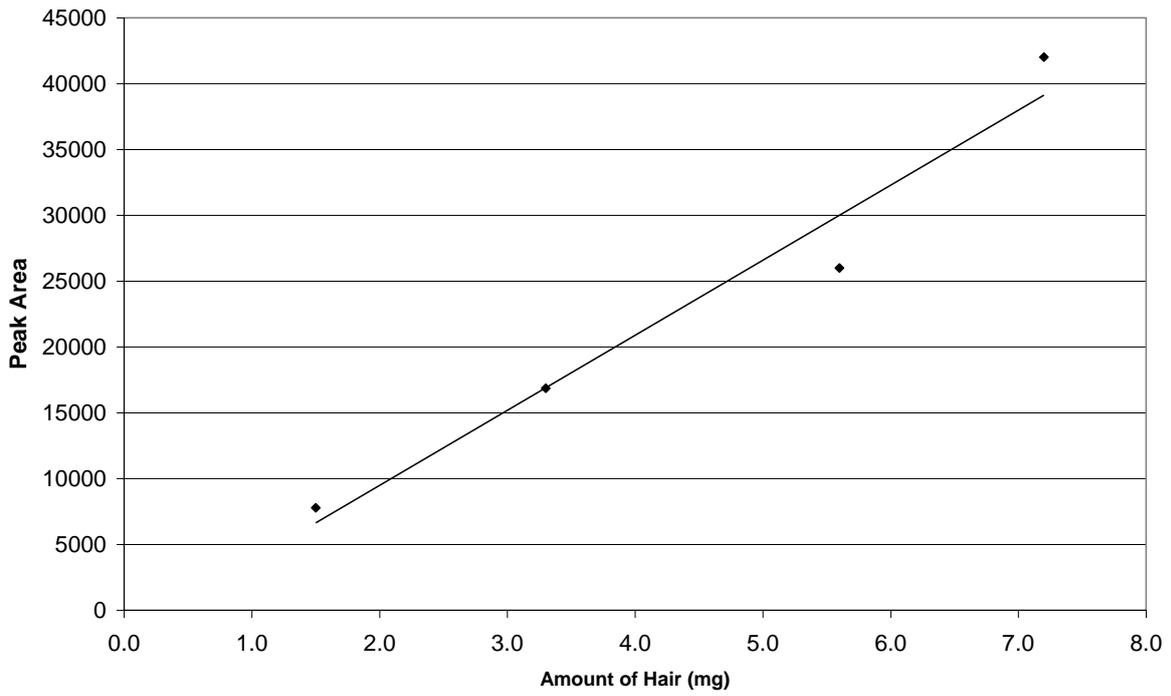


Figure 6.7 Results of the test for linearity of PTCA recovery.

6.3.2 PTCA content of Hair Samples

The raw PTCA values and colour data for each individual can be found in Appendix 1.

6.3.3 A Comparison of PTCA Content to Colour Measured by Reflective Spectrophotometry

Figure 6.8 shows the relationship between the macroscopic L^* value as measured by reflective spectrophotometry and the amount of PTCA recovered per milligram of hair. There was an expected negative correlation that indicates as hair gets lighter, there is less PTCA recovered and therefore, was less melanin present. This relationship was quite strong; however, very light hair samples ($L^* > 45$) and white hair samples were outside the main group. Non-European samples were also outside the main group, having much more PTCA per milligram of hair.

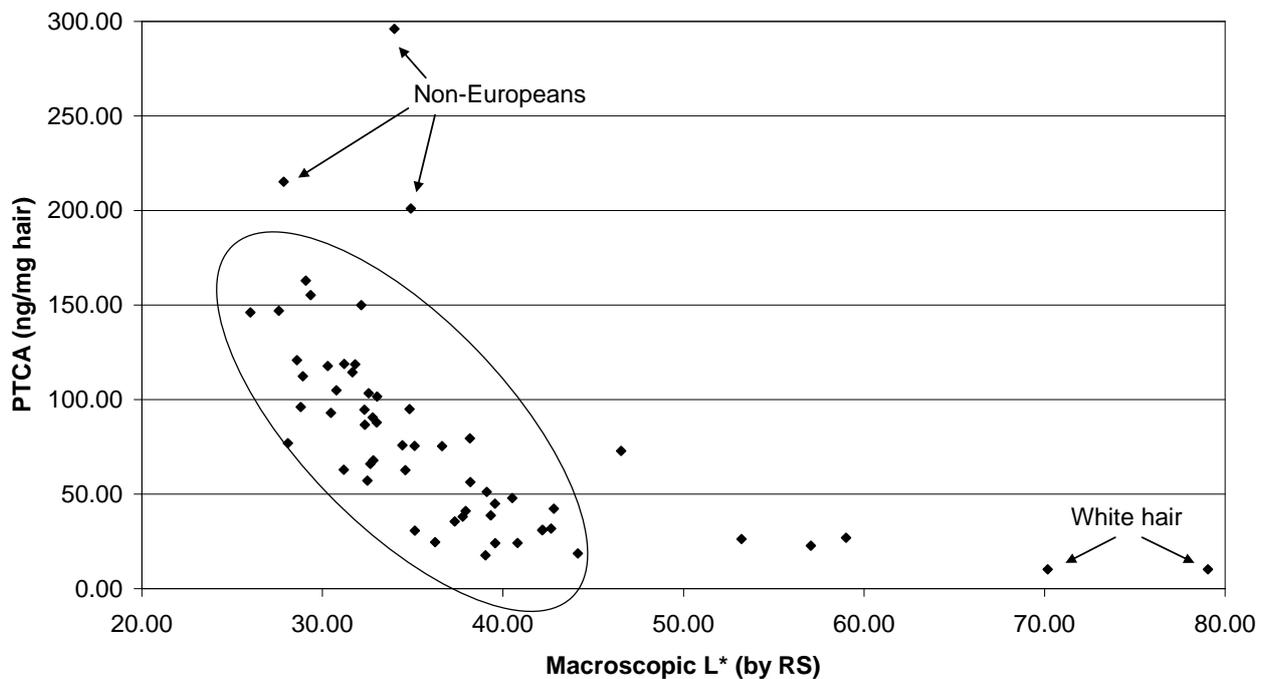


Figure 6.8 The amount of PTCA found in 1 mg of hair compared to the RS L^* value. The R^2 for the linear relationship of the entire sample set is 0.321 and the R^2 for the circled group is 0.6889. (p-values = 0.000)

The strongest linear relationship between the macroscopic colour of hair, for all samples, as measured by reflective spectrophotometry and the amount of PTCA recovered was along the b* axes, as seen in Figure 6.9. No significant relationship was found between the RS a* value and the amount of PTCA recovered ($R^2 = 0.185$, p-value = 0.094).

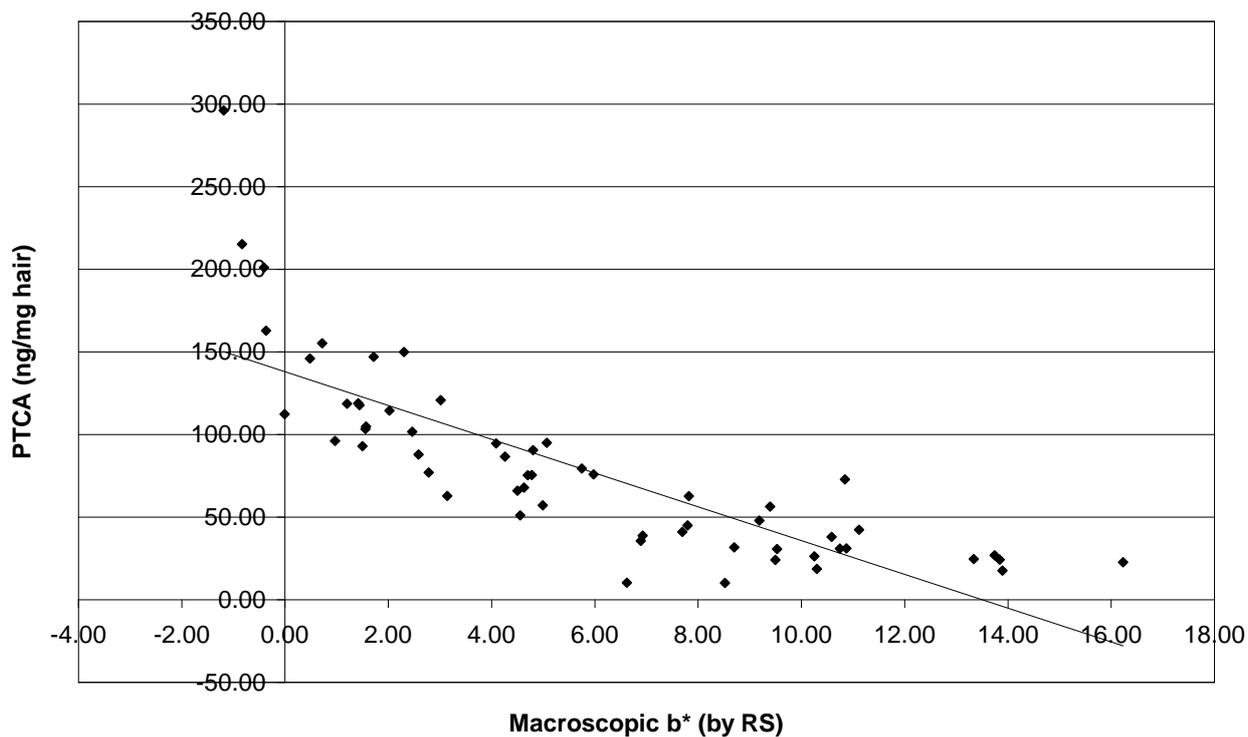


Figure 6.9 The amount of PTCA found in 1 mg of hair compared to the RS b* value ($R^2 = 0.6561$, p-value = 0.000)

6.3.4 A Comparison of PTCA Content to the Microscopic Colour

The strongest relationship between the microscopic colour and the amount of PTCA recovered was along the L* axis. Figure 4 illustrates this negative correlation. The microscopic a* value had a moderate positive correlation with the amount of PTCA recovered ($R^2 = 0.262$; p-value = 0.001), while the microscopic b* value showed no significant correlation.

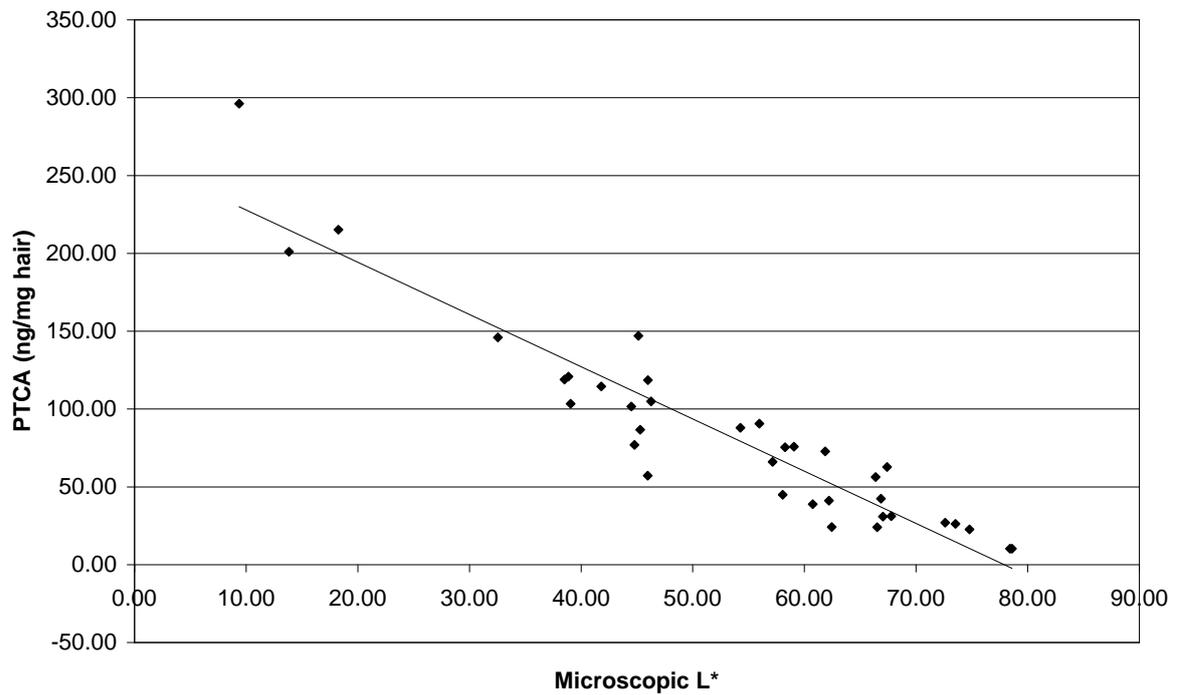


Figure 6.10 The amount of PTCA found in 1mg of hair compared to the Mc L* value ($R^2 = 0.8769$, p-value = 0.000)

6.3.5 A Comparison of PTCA Content to Hair Thickness

A moderate significant relationship was found between the amount of PTCA recovered and the thickness of the hair as measured in Chapter 5. This relationship can be seen in Figure 6.11.

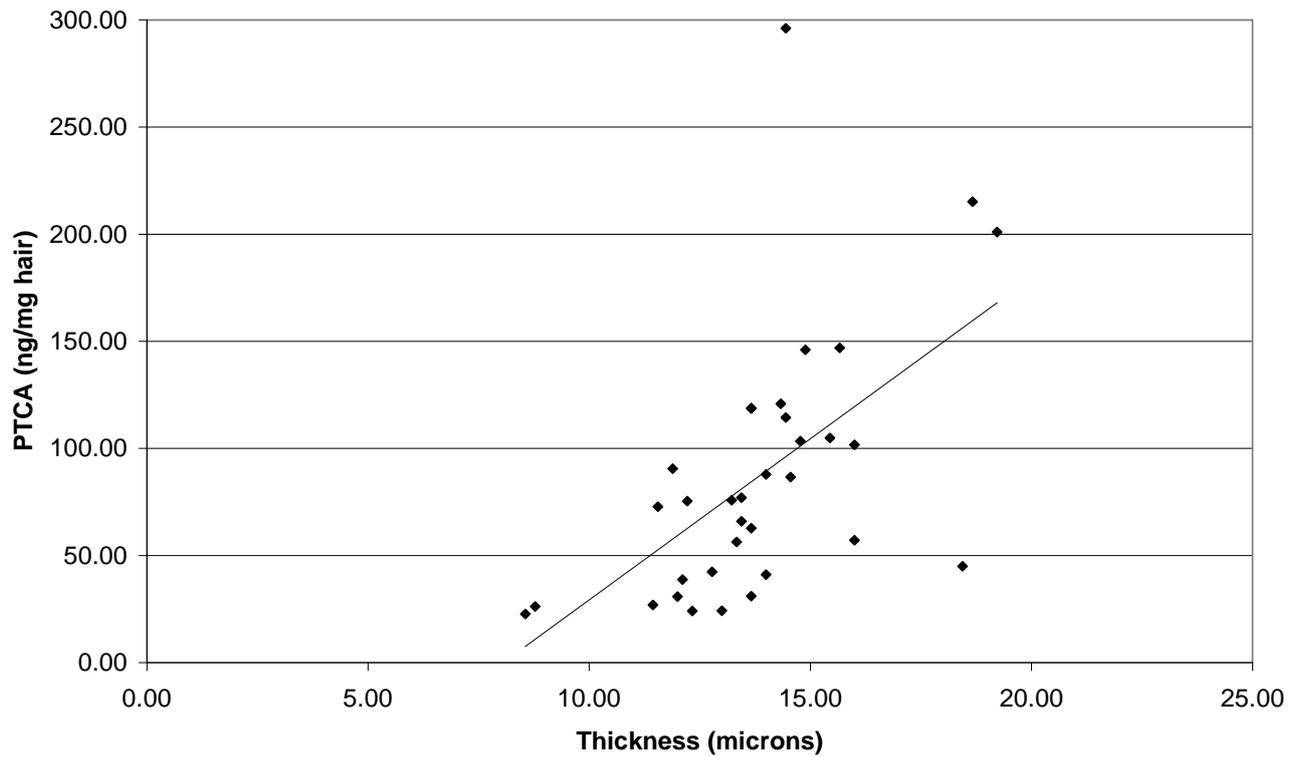


Figure 6.11 The mean thickness of the nine hair measurements compared to the amount of PTCA recovered from 1mg of hair ($R^2 = 0.323$, p-value = 0.000)

6.3.6 PTCA Content of Reflective Spectrophotometry Clusters

Chapter 3 examined hair colour measurements and found that those measurements may be broken into objective, describable clusters. The log-likelihood method was used to break 202 European individuals into five clusters (referred to as LL5); details of these clusters can be found in Chapter 2. The 58 individuals included in this chemical analysis were also analysed according to the cluster they were assigned to in Chapter 2. The mean PTCA content (ng PTCA/mg hair) of each cluster is shown in Figure 6.12, with the standard deviation represented by the error bars. In Figure 6.12, individuals of non-European ancestry are included for comparison and are referred to as Cluster 6, although they were not part of the LL5 analysis.

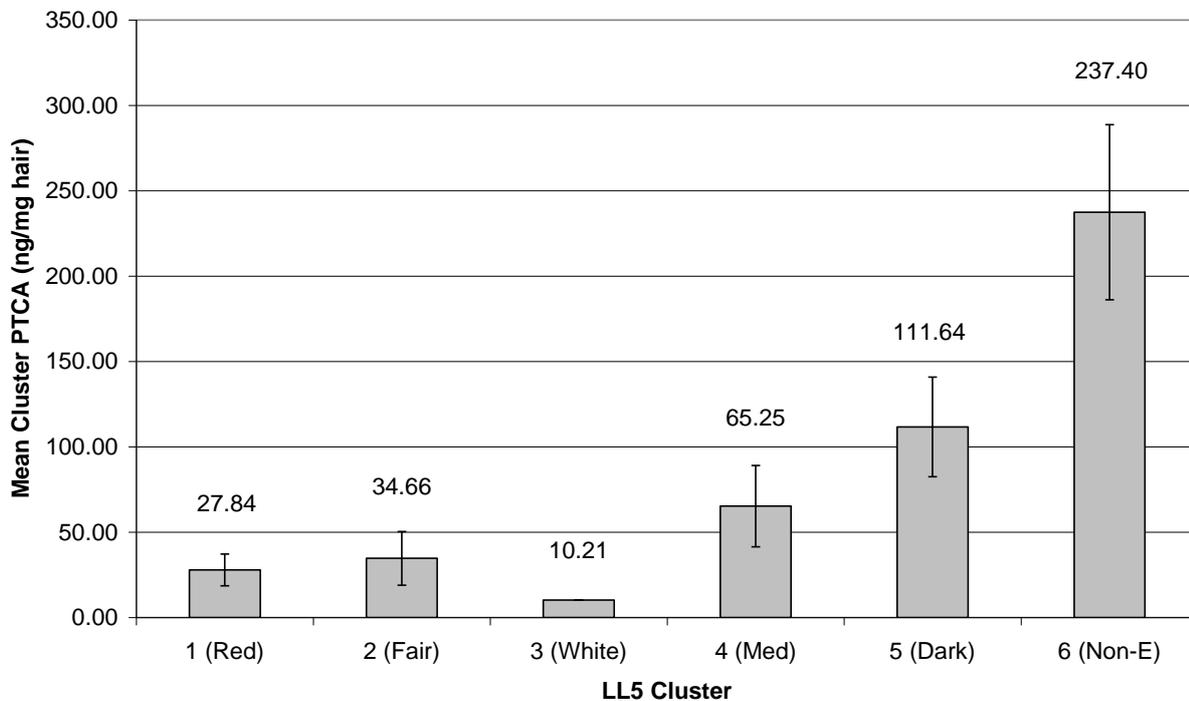


Figure 6.12 The mean content of PTCA for the LL5 reflective spectrophotometry clusters. Non-Europeans are referred to as Cluster 6 (but were not part of the LL5 analysis). The error bars represent the standard deviation in that cluster.

6.4 Discussion

The results of this experiment show how the amount of eumelanin in hair is related to the colour as measured in the CIE L*a*b* colour space by reflective spectrophotometry and microscopic analysis. As would be expected, lighter hair has less eumelanin; however, there is a stronger relationship between the amount of eumelanin present and the value of the b* (yellow) axis for macroscopic hair colour. The amount of PTCA measured is more closely related to the microscopic colour (specifically the L* axis) than to the macroscopic colour, which may be due to the different methods of colour measurement (reflective spectrophotometry compared to a digital microscopic image). Variation may also result from a chemical analysis being performed on cut ends of hair, where reflective spectrophotometry measurements were taken higher up on the head and microscopic measurements were taken along the length of the hair.

The PTCA results obtained for the clusters (Section 6.3.6) are also as expected. Clusters 5 (dark), 4 (Medium) and 2 (Fair) have progressively less of the eumelanin degradation product and all samples of European ancestry have a great deal less PTCA per milligram of hair than the non-European samples do. These results correspond reasonably well to previously published PTCA or Eumelanin results. Borges *et al.* (2001) also had approximately 300ng PTCA/mg hair for Black Asian hair [20], which corresponds to the non-European group in Section 6.3.6 and Napolitano *et al.* (2000) obtained PTCA results including the following: Black 264 ng/mg hair, Blonde 75 ng/mg hair, Light Blonde 55 ng/mg hair, Albino 10 ng/mg hair [82]. The origin of black hair in this case is unknown and the Blonde and Light Blonde are higher than the results obtained in this study, which may be a result of variation in methods, instrumentation or in the assignment of colour categories. It is interesting to note that the results for Type I Albino hair (Tyrosinase-negative) in Napolitano *et al.* (2000) were almost the same as those for White hair in this study – approximately 10ng PTCA per milligram of hair.

While the relationship between colour and eumelanin content of the hair is obvious, it is not perfect. Some variation in quantification may be responsible, but it may also be the case that there are other contributors to hair colour, such as the structure and thickness of the keratin matrix of the hair shaft, the distribution of the eumelanin pigment in the

hair and the amount and ratio of the pheomelanin pigment. Although the pheomelanin component of these samples was not quantified, the thickness of the hair was measured and was found to be moderately related to the amount of PTCA recovered. Two hairs with the same amount of eumelanin per milligram of hair may appear to be a different colour if the hair is thinner and the pigment is, therefore, more densely packed or is distributed in a different way.

While it is unfortunate that the quantification of pheomelanin content of these samples was unsuccessful and further work in this area would be beneficial, it should be noted that this study is primarily concerned with the characterisation of hair along the dark-fair continuum, where pheomelanin is a very minor component (approximately 5% of the pigment content [20]) and the ratio between eumelanin and pheomelanin less important than if one were studying red hair.

Chapter 7 Genetic Analysis

7.1 Introduction

7.1.1 Genetic Control of Pigmentation

It has been suggested that most pigmentation variation is probably the result of a few highly penetrant alleles [10], however, the search for these alleles and the genes they are hidden in is far from over. From studies of epidermal and iris pigmentation, as well as from murine models, we know some of the genes involved, or likely to be involved, in human pigmentation. Protein-coding genes TYR [86], TYRP 1 and 2 [86, 87], the P gene (OCA2) [86, 88, 89], MC1R [4, 14, 18, 86, 90], ASIP [40, 44], MATP [45], POMC [91], SILV [51, 86], MYO5A and RAB27A [5, 86], as well as transcription factors such as MITF or SOX10 have been implicated in human pigmentation [3, 6, 51].

Nonsynonymous mutations may lead to peptides with greater or lesser activity. Tyrosinase, the rate-limiting enzyme in melanin synthesis, has been found to have at least ten times greater activity in melanocytes derived from ‘black’ skin compared to those derived from ‘white’ skin [86], while the amount of protein remained the same. Normal variation of human pigmentation may also be the result of differences in regulatory elements at the transcriptional, translational or posttranslational levels [40], rather than in coding differences. For example, a SNP in the 3’ untranslated region of ASIP is correlated with darker pigmentation [44]. So far, mutations in the melanocortin-1 receptor have been the only variations directly correlated with hair colour differences; specifically red hair [4, 6, 18, 86, 90].

The following tables (Table 7.1 to Table 7.4) are adapted from the Albinism Database [92] and other relevant studies (cited in square brackets in the table) and indicate genes thought to be most involved in normal human pigmentation variation [92, 93] due to their disease-state (the appearance of individuals with a defective gene). Also listed are the chromosomal locations of the gene, the human disease or phenotype associated with a non-functioning protein, or a protein with reduced function, and the function or

potential function of the protein in normal pigmentation. There are likely to be other genes involved in different aspects of pigmentation (candidate pigmentation genes), such as *MLPH* [94] or *SLC24A5* [95]; genes that are logically connected to the pigmentation pathway but do not have phenotypes as obvious as those in Tables 7.1 to 7.4. Several of the genes tested in this study are of this type.

Table 7.1 Genes involved in the synthesis of melanin.

Human Locus	Alternate Name	Human Chromosome	Associated Disease	Protein Encoded	Function in Pigmentation
<i>TYR</i> [86]		11q21	<u>oculocutaneous albinism type 1 (OCA1)</u>	tyrosinase	Oxidises tyrosine and dopa to dopaquinone and polymerises DHI units into eumelanin
<i>TYRP1</i> [86, 87]		9p23	<u>oculocutaneous albinism type 3 (OCA3)</u> <u>Rufous albinism</u>	tyrosinase related protein 1 (TRP1)	Part of the tyrosinase complex (stabilising factor) and DHICA oxidase; incorporating DHICA units into eumelanin
<i>TYRP2</i> [86, 87]	<i>DCT</i>	13q31-q32	Unknown	tyrosinase related protein 2 (TRP2)	DOPAchrome tautomerase – modifying dopachrome to DHICA units
<i>MATP</i> [45]		5p	<u>oculocutaneous albinism type 4 (OCA4)</u>	Membrane-associated transporter protein	transporter
<i>Pmel-17</i> [51, 86]	<i>SILV</i> , <i>gp100</i>	12q13-q14	Unknown	gp100/silver protein, Pmel17	melanosomal matrix protein – specific marker for melanocytes

Table 7.2 Genes involved in melanosome structure and transport.

Human Locus	Alternate Name	Human Chromosome	Associated Disease	Protein Encoded	Function in Pigmentation
<i>MYO5A</i> [5, 86]		15q21	<u>Griscelli syndrome</u>	myosin type Va	Movement of the melanosome
<i>RAB27A</i> [5, 86]		15q21	<u>Griscelli syndrome</u>	RAS associated protein	Movement of the melanosome
<i>P</i> [86, 88, 89]	<i>OCA2</i>	15q11.2-q12	<u>oculocutaneous albinism type 2 (OCA2)</u>	Unkonwn; on melanosome membrane	Many proposed; ion exchange, pH control. Main gene involved in iris pigmentation

Table 7.3 Genes involved in the type of melanin produced (Eumelanin/Pheomelanin switch).

Human Locus	Human Chromosome	Associated Phenotype	Protein Encoded	Function in Pigmentation
<u>MC1R</u> [4, 14, 86, 90]	16q24.3	<u>hair colour/skin type</u>	Melanocortin 1 receptor	Eumelanin / pheomelanin switch
<u>ASIP</u> [40, 44]	20q11.2	<u>hair colour/skin type</u>	agouti signal protein (ASIP)	Eumelanin / pheomelanin switch
<u>POMC</u> [91]	2p23.3	Red Hair	Proopiomelanocortin; including Melanocyte Stimulating Hormone (MSH) Ligand to MC1R	Eumelanin / pheomelanin switch (and endocrine functions)

Table 7.4 Transcription factors that may be involved in human pigmentation.

Human Locus	Human Chromosome	Associated Phenotype	Protein Encoded	Function in Pigmentation
<u>MITF</u> [3, 6, 51]	3p12-14	<u>Waardenburg syndrome type 2</u>	transcription factor	melanocyte differentiation
<u>SOX10</u> [3, 6, 51]	22q13	<u>Waardenburg-Shah Syndrome</u>	SRY-box containing gene 10	transcriptional activator

Researchers are beginning to realise that there are fundamental differences between the genetic control of mouse and human pigmentation and between the pigmentation of human epidermis, iris and hair. It has recently been shown that Tyrosinase-Related Protein Type 2 (TYRP 2) is not expressed in hair follicle melanocytes found in the hair bulb, the site of melanogenesis for hair pigmentation [50]. Another study [91] has shown that there are major differences in the expression of the POMC/MC1R signalling pathway in hair follicle melanocytes, compared to epidermal melanocytes. In addition, several studies have shown that MC1R interacts with other genes, especially the P-gene [7, 15, 96] and possibly ASIP [40] and is not solely responsible for all hair colour variation [4, 40]. Much work needs to be done to elucidate the genetic contributors to normal hair colour variation.

7.1.2 Genetic Association Studies of Hair Colour

Sections 2.3.2 and 2.3.3 have discussed the results of and issues with genetic studies of hair colour. Like this investigation, these studies are typically candidate-gene association studies; where single nucleotide polymorphisms are examined in genes that have a logical connection to the pigmentation pathway and SNPs found to have significant associations may be mutations that cause variation in the resulting pigmentation, or they may be in linkage disequilibrium with mutations that do.

Population structure can lead to false positive associations when particular SNPs are ancestry-informative rather than phenotype specific [9]. For example, a SNP that has nothing to do with pigmentation may be associated with African ancestry and if there are persons of African ancestry in the sample, that SNP may end up being positively associated with darker hair, since people with African ancestry often have darker hair.

Choosing individuals with European ancestry limits the possibility of population structure interfering in the results of the genetic analysis.

7.1.3 Aims

It is hypothesised that there are genetic variants specifically associated with an objectively defined Blonde phenotype. Chapter 3 has illustrated one way of grouping hair colours into “clusters” according to their colour as measured by reflective spectrophotometry.

A total of 58 Single Nucleotide Polymorphisms (SNPs) were genotyped on 190 individuals of European ancestry and this Chapter will report on preliminary associations found with the clusters determined in Chapter 3 as well as quantitative measurements of CIE L^* , a^* and b^* . It is recognised that this is a small sample, but it is well-characterised with objective measurements and will provide an indication of which genes and polymorphism should be further considered. While there are many additional tests that could be conducted, the following are focused on associations with fair/blonde hair in order to best achieve the study aims.

7.2 Materials and Methods

7.2.1 Sampling

The genetic analysis included 190 individuals of European ancestry that volunteered a DNA sample, from the group described in 3.2.1. The two white hair individuals were not included. DNA samples were collected via buccal swab (four per individual) using either Omniswabs (Fitzco, Spring Park, MN) or cotton swabs (Copan Italia, Brescia, Italy). Omniswab heads were ejected into microcentrifuge tubes upon collection, allowed to dry and the tubes shut. Cotton swabs were placed in their sheaths with the ends cut off to allow drying and put in a sealed envelope.

7.2.2 Single Nucleotide Polymorphisms

The SNPs tested in this analysis are listed in Table 7.5 according to the SNP_ID number that can be used to find information about each polymorphism in the Entrez SNP database [97]. SNPs were chosen by Angela van Daal's research group at Bond University (Queensland, Australia) for a previous study of pigmentation associations and also covered most areas of interest for this research [98]. These SNPs were selected to cover candidate-pigmentation genes [99]. Candidate-pigmentation genes are those that can be connected to the melanin production pathway, such as those in the melanin synthesis pathway (e.g. TYR, MATP) or genes involved in melanosome structure (e.g. OCA2).

Table 7.5 Single Nucleotide Polymorphisms tested for association with hair colour and the genes in which they are found.

SNP_ID	Gene	SNP_ID	Gene
RS6453373	AP3B1	RS1805007	MC1R
RS1801244	ATP7B	RS885479	MC1R
RS1801243	ATP7B	RS1805005	MC1R
RS7334118	ATP7B	RS2228479	MC1R
RS1801249	ATP7B	RS4468625	MGRN1
RS732774	ATP7B	RS3751109	MLPH
RS17435026	ATP7B	RS3817362	MLPH
RS6107308	ATRN	RS2292884	MLPH
RS3886999	ATRN	RS3751107	MLPH
RS17782078	ATRN	RS11883500	MLPH
RS16876571	DTNBP1	RS6435927	MREG
RS17470454	DTNBP1	RS1058219	MYO5A
RS17336807	Egfr	RS16964944	MYO5A
RS11543848	Egfr	RS1724577	MYO5A
RS2296434	HPS1	RS1052030	MYO7A
RS2296436	HPS1	RS2276288	MYO7A
RS3747129	HPS4	RS7117511	MYO7A
RS2014410	HPS4	RS2305253	OCA2
RS1894706	HPS4	RS1800414	P(OCA2)
RS1894704	HPS4	RS28945095	PAX3
RS5752330	HPS4	RS729421	RABGGTA
RS7128017	HPS5	RS17118154	SILV (Pmel-17)
RS14024	KRT1	RS1426654	SLC24A5
RS2634041	KRT2	RS26722	SLC45A2 (MATP)
RS638043	KRT2	RS16891982	SLC45A2 (MATP)
RS7541041	LYST	RS1800422	TYR
RS3768067	LYST	RS1042602	TYR
RS10465613	LYST	RS16929374	TYRP1
RS7524261	LYST		

7.2.3 DNA Extraction, Quantification, Preparation and Genotyping

DNA was extracted from two buccal swabs from each individual using the Qiagen Genra Puregene Buccal Cell Kit according to the manufacturer's instructions with some modification. The two swab heads were placed into one microcentrifuge tube (cotton swabs had the heads cut off the wooden shaft). 600 μ L of Qiagen cell lysis solution was used with cotton swabs and 800 μ L with Omniswabs to completely cover both swab heads. 3 μ L of Proteinase K (>600 mAU/mL; Qiagen) and RNase A (17500U/mL;Qiagen) was used with the cotton swabs and 4 μ L with the Omniswabs. One mAU is the activity that releases folin-positive amino acids and peptides corresponding to 1 μ mol tyrosine per minute [100]. 200 μ L of Protein Precipitation solution were used for both swab types. 600 μ L of isopropanol was used to precipitate DNA according to the kit procedure and samples were hydrated in 40 μ L of sterile, distilled and deionised water.

DNA samples were quantified using the Quantifiler Human DNA Quantification kit (Applied Biosystems Australia, Victoria, Australia), a real-time qPCR method which was performed according to standard protocols at the Victoria Police Forensic Services Department [101, 102].

If samples had more than 10ng/ μ L of DNA, they were diluted to this concentration using sterile ddH₂O. If the concentration was between 5ng/ μ L and 10ng/ μ L, the samples were concentrated using an isopropanol DNA precipitation [103] as follows; 1 Volume of DNA sample was added to 0.1 Volume of Sodium Acetate and 1 Volume of isopropanol and held at -20°C overnight. The sample was centrifuged at 14 000g and the supernatant removed. The precipitated DNA pellet was washed with 70% ethanol and dried. The sample was then re-hydrated in the appropriate amount of sterile ddH₂O to make 10ng/ μ L (assuming 100% recovery). If samples had less than 5ng/ μ L of DNA, the last two swabs were extracted and the procedure conducted again. The above procedures resulted in successful DNA recovery for each individual.

10 μ L of each DNA sample with a concentration of 10ng/ μ L were loaded into 96-well plates and were sent to the Australian Genome Research Facility for genotyping via the

Sequenom MassARRAY system. A total of 58 SNPs were genotyped, of which, 46 were found to be polymorphic in this population.

7.2.4 Software and Statistical Analysis

Genetic analysis was performed with software by Golden Helix (Bozeman, MT); HelixTree and the SNP Variation Suite (free trial). Allele frequency data was analysed using Chi squared tests (χ^2) and genotype data was analysed using Fisher's Exact Test. Hardy-Weinberg Equilibrium and linkage disequilibrium were also assessed using this software. Regression analysis to test for second-order interaction terms was also performed with the HelixTree software.

No corrections were made for multiple analyses and the results, therefore, may reflect some degree of false positive associations. Bonferroni adjustments eliminated any significant associations, but this method is thought to be overly conservative for this study [104, 105]. Previous hair colour genetics studies do not mention the use of multiple analysis correction [38, 45].

Due to the small sample size in this study, $\alpha = 0.1$ was used as a relaxed requirement of significance.

7.3 Results

7.3.1 Polymorphic Single Nucleotide Polymorphisms

Of the 58 SNPs tested, 46 were found to be polymorphic. Two SNPs, rs17336807 and rs11543848 are recognised as the same polymorphism which is now referred to as rs2227983 [106]. Eleven SNPs were found to be non-polymorphic. All of these were expected and have been reported as being fixed in European populations by the International HapMap project [107]. However, one SNP that has been reported as fixed in Europeans by HapMap was found to be polymorphic in this sample. The SNP rs1426654 had a minor allele frequency of 0.014 in this sample, which is consistent with other research [95].

Table 7.6 shows the minor allele frequency of the typed SNPs as well as the call (or success) rate of genotyping. Raw genotype data may be found in Appendix 5.

Table 7.6 SNP results showing the minor allele frequency and the call rate – non-polymorphic SNPs have a minor allele frequency of 0.00.

Marker	Minor Allele Freq.	Call Rate	Marker	Minor Allele Freq.	Call Rate
RS1042602	0.339	0.979	RS2228479	0.067	0.984
RS10465613	0.000	0.995	RS2276288	0.426	0.989
RS1052030	0.353	0.984	RS2292884	0.213	0.989
RS1058219	0.169	0.979	RS2296434	0.080	0.984
RS1061472	0.428	0.989	RS2296436	0.080	0.984
RS11543848	0.293	0.968	RS2305253	0.000	0.995
RS11883500	0.147	0.968	RS2634041	0.000	0.995
RS14024	0.297	0.984	RS26722	0.021	0.984
RS1426654	0.014	0.968	RS28945095	0.005	0.984
RS16876571	0.013	0.995	RS3747129	0.144	0.984
RS16891982	0.051	0.989	RS3751107	0.141	0.968
RS16929374	0.000	0.995	RS3751109	0.144	0.989
RS16964944	0.000	0.995	RS3768067	0.000	0.995
RS17118154	0.003	0.984	RS3817362	0.091	0.863
RS1724577	0.000	0.995	RS3886999	0.080	0.989
RS17336807	0.293	0.989	RS4468625	0.000	0.995
RS17435026	0.429	0.958	RS5752330	0.130	0.989
RS17470454	0.059	0.989	RS6107308	0.005	0.989
RS17782078	0.080	0.989	RS638043	0.147	0.968
RS1800414	0.000	0.995	RS6435927	0.212	0.979
RS1800422	0.340	0.984	RS6453373	0.067	0.984
RS1801243	0.487	0.984	RS7117511	0.003	0.989
RS1801244	0.484	0.984	RS7128017	0.120	0.968
RS1801249	0.420	0.989	RS729421	0.385	0.984
RS1805005	0.133	0.968	RS732774	0.420	0.989
RS1805007	0.110	0.984	RS7334118	0.016	0.984
RS1894704	0.090	0.989	RS7524261	0.000	0.995
RS1894706	0.090	0.989	RS7541041	0.000	0.995
RS2014410	0.398	0.984	RS885479	0.055	0.963

7.3.2 Hardy-Weinberg Equilibrium and Linkage Disequilibrium

All SNPs except three were in Hardy-Weinberg Equilibrium, which may be due to the effects of natural selection or non-random mating. SNPs not in HWE were rs6453373, rs16891982 and rs17435026.

Figure 7.1 illustrates the low amount of linkage disequilibrium found amongst the markers used in this study (blue means a low amount of LD). The plot takes into account which markers are on the same chromosome and which may be linked.

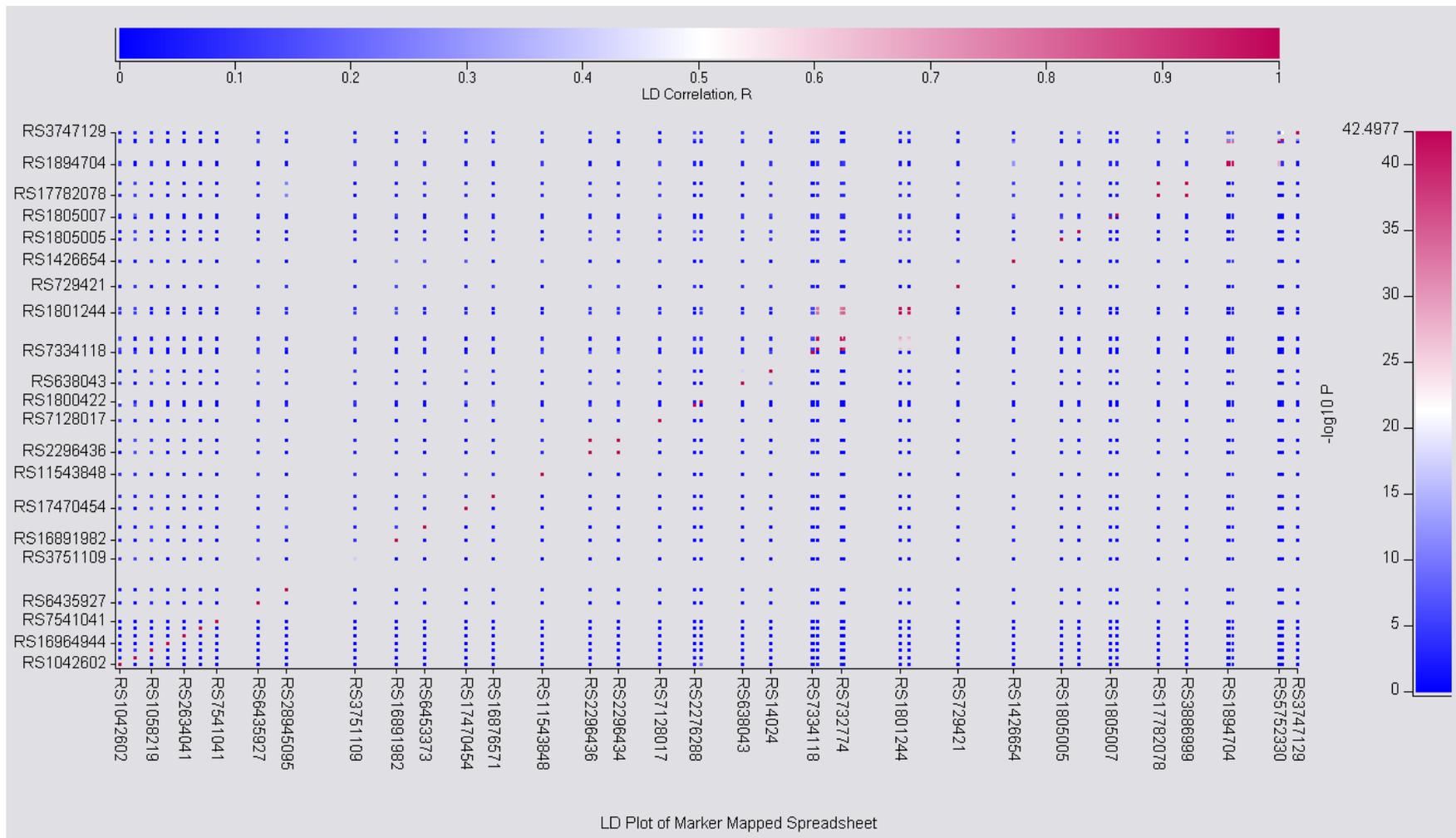


Figure 7.1 Linkage disequilibrium plot of SNP markers that have been mapped according to their chromosomal locations.

7.3.3 Tests for Associations with Categorical Variables (Clusters)

Several genetic models were tested for SNPs that were polymorphic in this population. Individuals in the sample were treated as *cases* or *controls*, with *cases* being the particular hair colour phenotype under study and *controls* being the rest of the population. The Odds Ratio (OR) is reported for the effect that the minor allele or particular genotype being considered has on the *case* phenotype. An odds ratio of 1 means that this allele or genotype is equally likely to be found in both cases and controls. An odds ratio greater than one means that this particular allele or genotype is more likely to be found in the cases than in the controls. An odds ratio less than one means that this particular allele or genotype is more likely to be found in the controls than in the cases (this is the case when the major allele is associated with the phenotype being considered).

The SNP being considered, the minor allele, the odds ratio, the gene that the SNP can be found in and the amino acid substitution it causes are reported for each of the models tested (Table 7.7 Table 7.15). In Tables 7.6 to 7.14; N equals the number of cases, D represents the minor allele and d represents the major allele.

The whole population was considered first. It is well known that variations in the gene MC1R contribute to Red hair colour [18], so the first allelic test used individuals in Cluster 1 (Red) as cases (Table 7.7). Table 7.8 shows the SNPs that were found to be significantly associated with Cluster 2 (Blonde/Fair) at the allelic level.

Table 7.7 SNPs found to be significantly associated with LL5 Red (Cluster 1) at the $\alpha = 0.1$ level for the allelic model (N = 8)

D vs. d				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs1805007	T	13.1	MC1R	R151C
rs1800422	A	3.4	TYR	R402Q

Table 7.8 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the allelic model (N = 32)

D vs. d				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs16876571	A	7.7	DTNBP1	H297Y
rs2276288	A	1.6	MYO7A	C1628S

The following tests use Cluster 2 (Fair/Blonde) as the cases for a dominant model of association (Table 7.9) and a recessive model (Table 7.10).

Table 7.9 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the dominant model (N = 32)

DD/Dd vs. dd				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs16876571	A	8.0	DTNBP1	H297Y
rs2276288	A	2.4	MYO7A	C1628S
rs1805007	T	2.1	MC1R	R151C

Table 7.10 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the recessive model (N = 32)

DD vs. Dd/dd				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs1800422	A	3.1	TYR	R402Q
rs17435026	A	2.1	ATP7B	R832K

Since the variations in MC1R leading to red hair are well known and can already be tested for [14, 18, 37-41, 108], additional tests were conducted with the eight individuals from Cluster 1 (Red) removed from the population. This was done to independently test for associations with Fair/Blonde hair colour at the allelic level (Table 7.11) and with a dominant (Table 7.12) or recessive model (Table 7.13).

Table 7.11 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the allelic model with Cluster 1 removed (N = 32)

D vs. d				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs16876571	A	7.3	DTNBP1	H297Y
rs1805007	T	2.3	MC1R	R151C
rs2276288	A	1.7	MYO7A	C1628S
rs1800422	A	1.7	TYR	R402Q

Table 7.12 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the dominant model with Cluster 1 removed (N = 32)

DD/Dd vs. dd				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs16876571	A	7.6	DTNBP1	H297Y
rs1805007	T	2.6	MC1R	R151C
rs2276288	A	2.4	MYO7A	C1628S

Table 7.13 SNPs found to be significantly associated with LL5 Fair/Blonde (Cluster 2) at the $\alpha = 0.1$ level for the recessive model with Cluster 1 removed (N = 32)

DD vs Dd/dd				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs1800422	A	3.8	TYR	R402Q

Tests for genetic associations with LL5 Cluster 5 (dark hair) were conducted to see if any SNPs had the effect of contributing to darker hair colours and may indicate an opposing process than that which contributes to light hair colours. Table 7.14 and Table 7.15 show the associations found in an allelic and a dominant model. Three SNPs showed associations with LL5 Cluster 5 (Dark) and all were different SNPs than had been associated with Red or Blonde/Fair hair colour (and not simply the opposite allele). However; two SNPs were in *ATP7B*, a gene that had previously shown a different SNP to be associated with Blonde/Fair hair (See Table 7.10).

Table 7.14 SNPs found to be significantly associated with LL5 Dark (Cluster 5) at the $\alpha = 0.1$ level for the allelic model (N = 86)

D vs d				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs16891982	C	3.86	MATP	L374F

Table 7.15 SNPs found to be significantly associated with LL5 Dark (Cluster 5) at the $\alpha = 0.1$ level for the dominant model (N = 86)

DD/Dd vs. dd				
SNP	Minor Allele	OR	Gene	Amino Acid Change
rs16891982	C	3.46	MATP	L374F
rs1801243	G	0.49	ATP7B	S406A
rs1801244	C	0.49	ATP7B	V456L

7.3.4 Tests for Associations with Quantitative Variables

Associations with the quantitative variables L, a* and b* were examined next. The mean of 15 reflective spectrophotometry measurements (as described in Section 3.2.2) was used to test for associations. A significance level of $\alpha = 0.1$ was used. The Corr/Trend (Correlation/Trend) is a linear regression test that will show the p-value for the dependent variable (L*, a* or b*) value having any correlation with or “trend” which depends upon the count value of the genotype. The Corr/Trend R value indicates the degree and direction of this trend. The F-test is also used, as this test is on whether the distributions of the dependent variable (L*, a* or b*) within each category are significantly different between the various categories of the predictor variable (genotypes DD, Dd and dd). These tests were performed on the whole sample; no individuals (like the Reds in 7.3.2) were removed.

Table 7.16 illustrates allelic associations with L*. It shows that five SNPs have a positive correlation with L*, meaning the minor allele is associated with lighter hair (higher L* values) while rs16891982 shows a negative correlation with L* (the minor allele is associated with darker hair).

Table 7.16 SNPs that have significant allelic associations with L* (D vs. d)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
rs16891982	0.005	-0.144	0.005	MATP	L374F
rs1800422	0.005	0.144	0.005	TYR	R402Q
rs1805007	0.029	0.113	0.029	MC1R	R151C
rs16876571	0.040	0.106	0.040	DTNBP1	H297Y
rs17782078	0.079	0.091	0.079	ATR1	I426T
rs3886999	0.079	0.091	0.079	ATR1	R1152K

The F-test statistic shown in Table 7.17 is used to test whether the distribution of the dependent variable (L^*) within each category are significantly different between the various possible genotypes (DD, Dd or dd). The SNPs shown in Table 7.17 have significant differences of L^* between the three possible genotypes.

Table 7.17 A test for genotypic associations with L^* (DD vs. Dd vs. dd)

SNP	F-Test P	Gene	Amino Acid Change
rs1800422	0.007	TYR	R402Q
rs16891982	0.035	MATP	L374F
rs16876571	0.039	DTNBP1	H297Y
rs3751107	0.041	MLPH	G172D
rs3751109	0.056	MLPH	L153P
rs11883500	0.062	MLPH	T289I
rs1052030	0.068	MYO7A	T16S

Table 7.18 and Table 7.19 show significant associations between various SNPs and L^* in a dominant and recessive model, respectively. As mentioned above, positive correlations (R) indicate that the minor allele is associated with higher (lighter) L^* values while negative correlations (R) indicate that the minor allele is associated with lower (darker) L^* values.

Table 7.18 SNPs that have a significant association with L^* in dominant model (DD, Dd vs dd)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
rs16891982	0.012	-0.184	0.012	MATP	L374F
rs1805007	0.036	0.153	0.036	MC1R	R151C
rs16876571	0.039	0.150	0.039	DTNBP1	H297Y
rs11883500	0.046	-0.148	0.045	MLPH	T289I
rs1800422	0.050	0.144	0.050	TYR	R402Q
rs3751107	0.064	-0.137	0.064	MLPH	G172D
rs3751109	0.084	-0.126	0.084	MLPH	L153P
rs1052030	0.089	0.125	0.089	MYO7A	T16S

Table 7.19 SNPs that have a significant association with L* in a recessive model (DD vs Dd, dd)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
rs1800422	0.004	0.212	0.004	TYR	R402Q
rs1058219	0.089	0.125	0.089	MYO5A	R1246C

Tables 7.17 and 7.18 show that rs1800422 is significantly associated with L* in both a dominant and recessive model. The model with the stronger trend (recessive in this case) is more likely to be the correct one. A more obvious example of this in MC1R is shown below (Figure 7.2 and Figure 7.3).

Next, associations with the quantitative variable a* were examined. Table 7.20 shows SNPs where the presence of the minor allele is significantly correlated with a*. Positive correlations indicate an association with hair becoming redder, while negative associations indicate an allelic association with hair that is less red.

Table 7.20 SNPs that have significant allelic associations with a* (D vs. d)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
rs1805007	0.000	0.265	0.000	MC1R	R151C
rs1800422	0.002	0.158	0.002	TYR	R402Q
rs16891982	0.033	-0.110	0.033	DTNBP1	H297Y
rs11883500	0.041	-0.107	0.041	MLPH	T289I
rs17470454	0.065	-0.095	0.065	DTNBP1	P272S
rs3751107	0.076	-0.092	0.076	MLPH	G172D
rs3751109	0.090	-0.088	0.090	MLPH	L153P

A test for the difference between the three genotypes was conducted. The SNPs shown in Table 7.21 have significant differences of a* between the three possible genotypes of DD, Dd and dd.

Table 7.21 A test for genotypic associations with a* (DD vs. Dd vs. dd)

SNP	F-Test P	Gene	Amino Acid Change
rs1805007	1.96E-10	MC1R	R151C
rs1800422	0.005	TYR	R402Q
rs3751107	0.056	MLPH	G172D
rs11883500	0.056	MLPH	T289I
rs3751109	0.072	MLPH	L153P

Table 7.22 and Table 7.23 show significant associations between various SNPs and a* in a dominant and recessive model, respectively. As mentioned above, positive correlations (R) indicate that the minor allele is associated with higher (more red) a* values while negative correlations (R) indicate that the minor allele is associated with lower (less red) a* values.

Table 7.22 SNPs that have a significant association with a* in dominant model (DD, Dd vs dd)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
rs1805007	0.000	0.277	0.000	MC1R	R151C
rs1800422	0.020	0.171	0.019	TYR	R402Q
rs11883500	0.021	-0.170	0.021	MLPH	T289I
rs3751107	0.032	-0.159	0.031	MLPH	G172D
rs3751109	0.040	-0.150	0.039	MLPH	L153P
rs16891982	0.077	-0.129	0.077	MATP	L374F
rs17470454	0.080	-0.128	0.080	DTNBP1	P272S

Table 7.23 SNPs that have a significant association with a* in a recessive model (DD vs Dd, dd)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
rs1805007	0.000	0.430534937	0.000	MC1R	R151C
rs1800422	0.004	0.209	0.004	TYR	R402Q

Tests with a^* show two SNPs that are significantly associated with a^* in both the dominant and recessive model. This happens when phenotypes are not perfectly associated with genotype; however the stronger correlation/trend is likely to be the correct one (recessive in both of these cases). The SNP from MC1R (rs1805007; R151C) is used as an example. The correlation of R151C with a^* is 0.277 in the dominant model and 0.43 in the recessive model. Figure 7.2 shows the slight trend that occurs if the minor allele (D) is treated as a dominant allele (0.277) while Figure 7.3 shows the stronger trend that occurs if D is treated as a recessive allele.

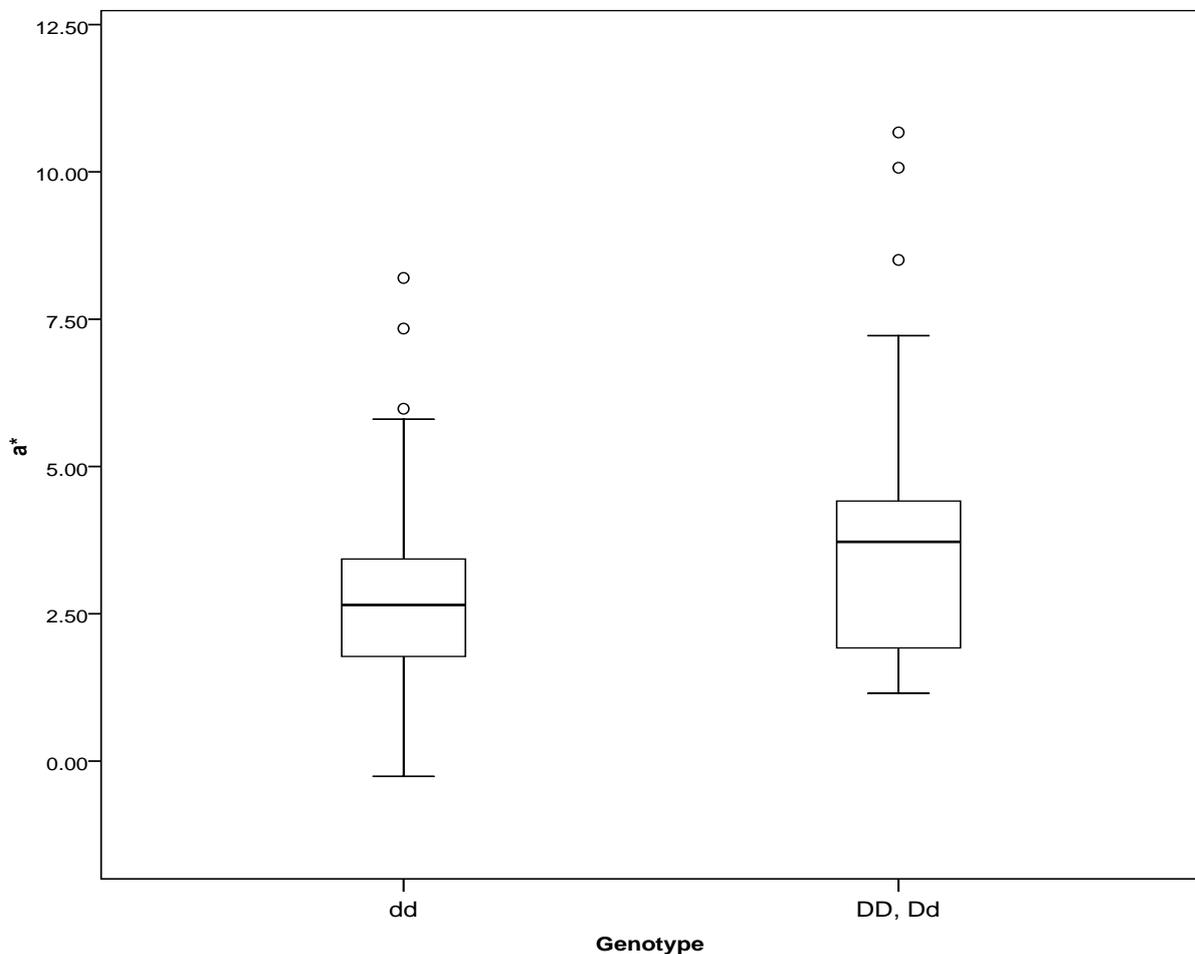


Figure 7.2 The dominant genotype model for association with a^* for rs1805007 (MC1R; R151C). D represents the minor allele and d represents the major allele.

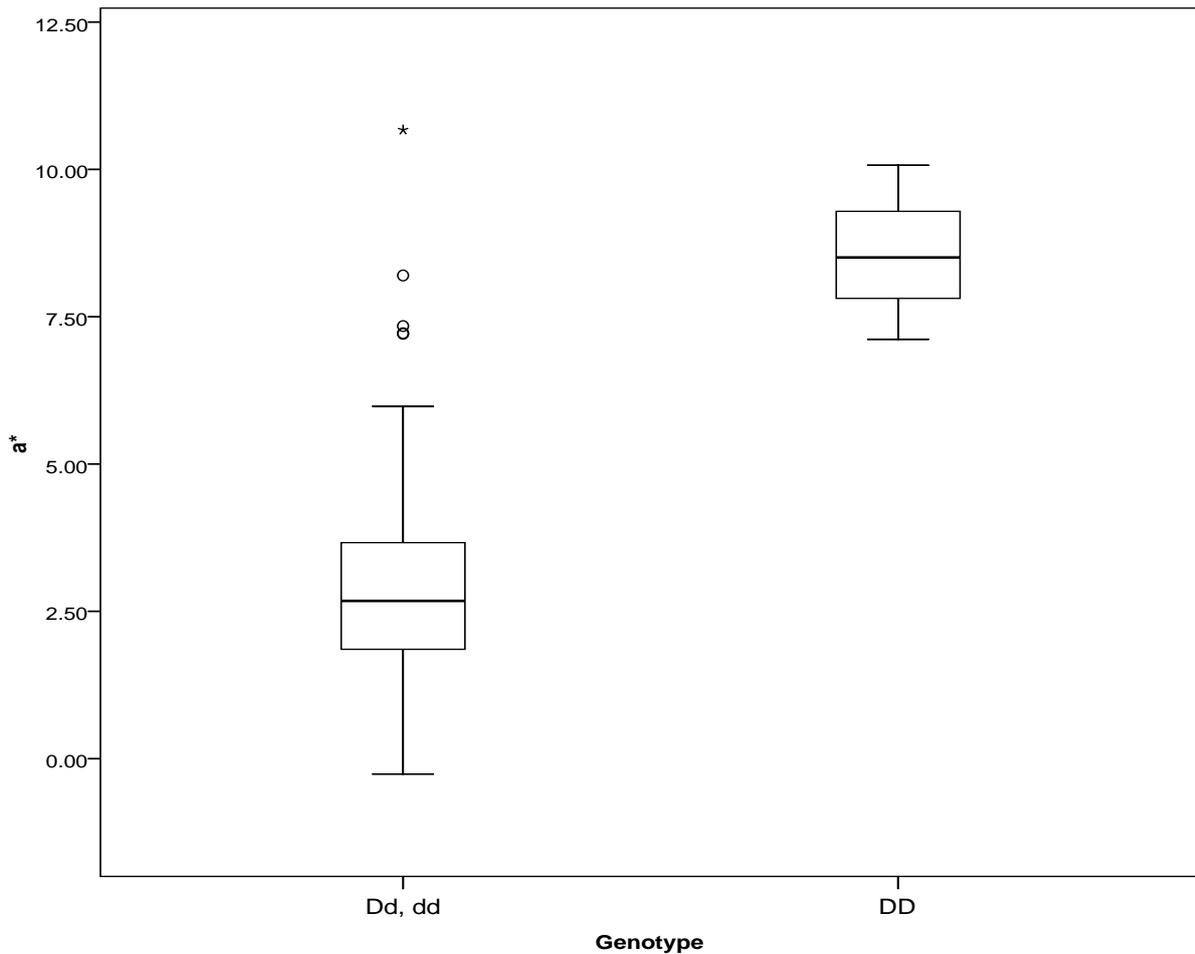


Figure 7.3 The recessive genotype model for association with a^* for rs1805007 (MC1R; R151C). D represents the minor allele and d represents the major allele.

Finally, associations with the quantitative variable b^* were examined. Table 7.24 shows SNPs where the presence of the minor allele is significantly correlated with b^* . Positive correlations indicate an association with hair becoming more yellow, while negative associations indicate an allelic association with hair that is less yellow.

Table 7.24 SNPs that have significant allelic associations with b* (D vs. d)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
rs1805007	0.000	0.214	0.000	MC1R	R151C
rs16891982	0.002	-0.162	0.002	MATP	L374F
rs1800422	0.004	0.148	0.004	TYR	R402Q
rs7334118	0.018	-0.122	0.018	ATP7B	H1207R
rs11883500	0.078	-0.092	0.078	MLPH	T289I

A test for the difference between the three genotypes was conducted for b* as well. The SNPs shown in Table 7.25 have significant differences of b* between the three possible genotypes of DD, Dd and dd.

Table 7.25 A test for genotypic associations with b* (DD vs. Dd vs. dd)

SNP	F-Test P	Gene	Amino Acid Change
rs1805007	0.000	MC1R	R151C
rs1800422	0.004	TYR	R402Q
rs16891982	0.012	MATP	L374F
rs7334118	0.017	ATP7B	H1207R
rs3751107	0.044	MLPH	G172D
rs11883500	0.064	MLPH	T289I
rs3751109	0.069	MLPH	L153P
rs729421	0.083	RABGGTA	A420T
rs1052030	0.086	MYO7A	T16S

Table 7.26 and Table 7.27 show significant associations between various SNPs and b* in a dominant and recessive model, respectively. As mentioned above, positive correlations (R) indicate that the minor allele is associated with higher (more yellow) b* values while negative correlations (R) indicate that the minor allele is associated with lower (less yellow) b* values.

Table 7.26 SNPs that have a significant association with b* in dominant model (DD, Dd vs dd)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
RS1805007	0.000	0.261	0.000	MC1R	R151C
RS16891982	0.003	-0.215	0.003	MATP	L374F
RS7334118	0.017	-0.174	0.017	ATP7B	H1207R
RS11883500	0.034	-0.157	0.033	MLPH	T289I
RS1800422	0.053	0.142	0.053	TYR	R402Q
RS3751107	0.055	-0.142	0.055	MLPH	G172D
RS3751109	0.081	-0.128	0.081	MLPH	L153P
RS729421	0.086	-0.126	0.086	RABGGTA	A420T
RS1052030	0.088	0.125	0.089	MYO7A	T16S

Table 7.27 SNPs that have a significant association with b* in a recessive model (DD vs Dd, dd)

SNP	Corr/Trend P	Corr/Trend R	F-Test P	Gene	Amino Acid Change
RS1805007	0.002	0.230	0.002	MC1R	R151C
RS1800422	0.002	0.227	0.002	TYR	R402Q
RS1058219	0.084	0.127	0.084	MYO5A	R1246C

7.3.5 Genetic Interactions

Genetic interactions contributing to Fair/Blonde hair were tested for by performing a regression analysis and looking for second order terms using LL5 Cluster 2 (Fair/Blonde) as the dependant variable (1= Cluster 2, 0 = Not Cluster 2). The resulting dihybrid genotypes were analysed with the HelixTree Software. Only one significant interaction was found within this sample; between rs16891982 in the MATP gene (L374F) and rs1805005 in MC1R (V60L). The figures below illustrate the significant associations found when the genotypes of rs16891982 and rs1805005 are considered together. Figure 7.4 shows the association of the dihybrid genotypes with Cluster 2 (Fair/Blonde hair) and it can be seen that individuals with genotype GC_GT are all in Cluster 2 (which has a value of 1)

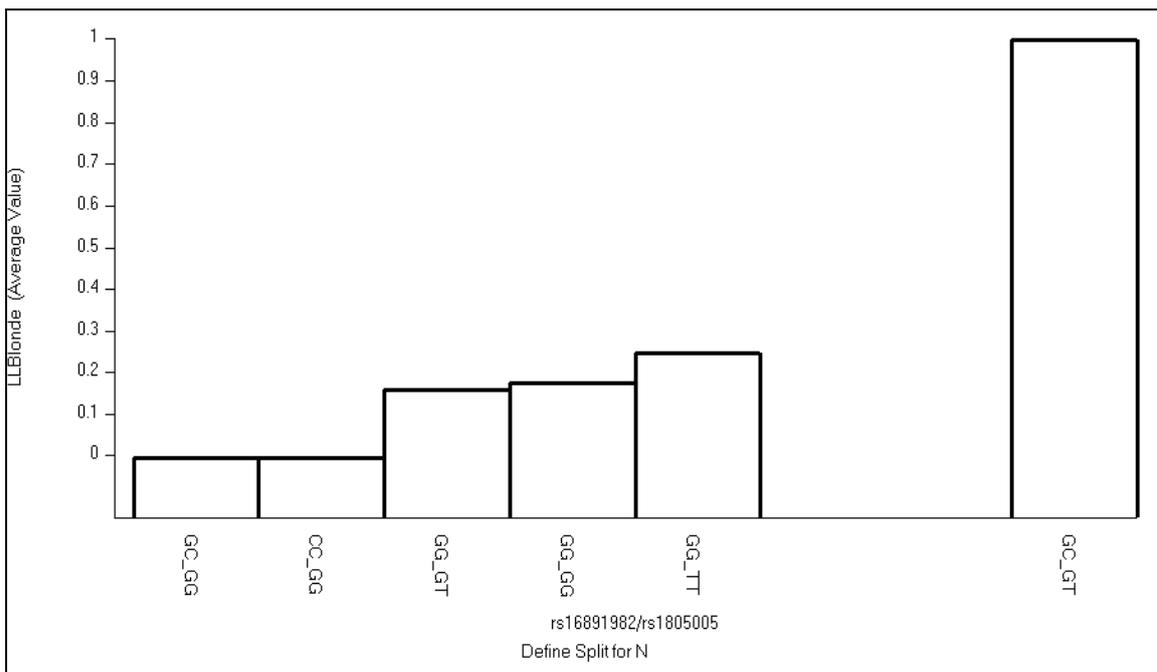


Figure 7.4 The associations between the rs16891982 and rs1805005 dihybrid and the LL5 Blonde/Fair Cluster 2.

Figure 7.5 and Figure 7.6 show the relationship between the dihybrid genotypes and the quantitative variables L* and b* (a* showed no significant relationship). It should be noted that the genotypes along the x-axis are not always in the same order.

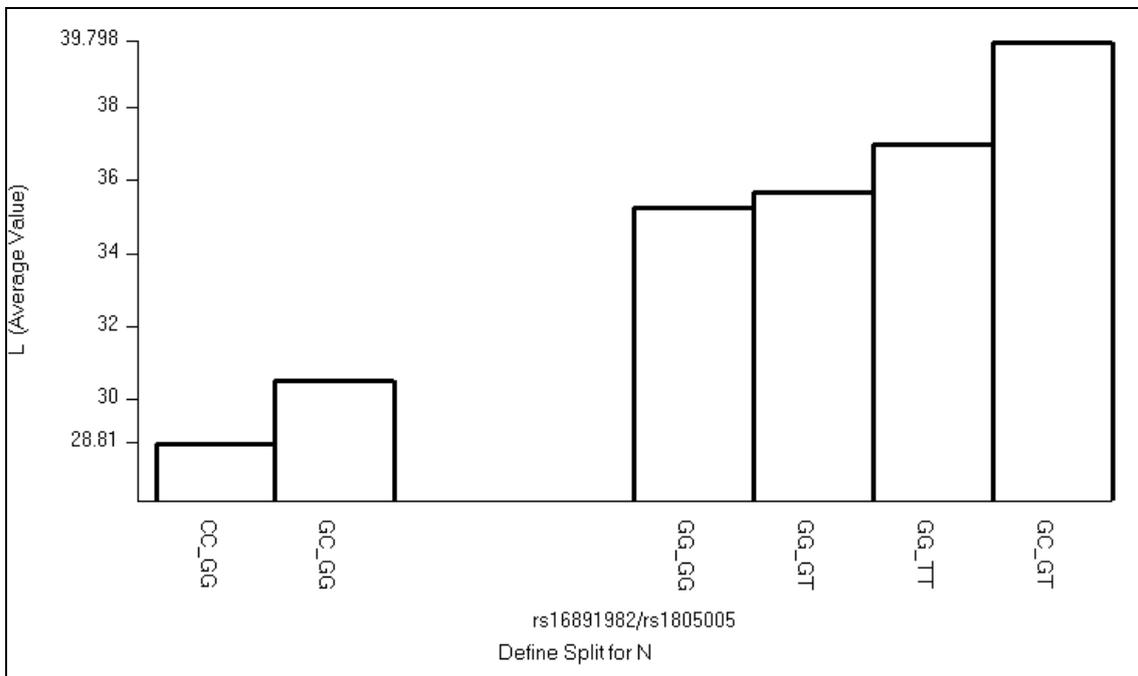


Figure 7.5 The associations between the rs16891982 and rs1805005 dihybrid and L* value

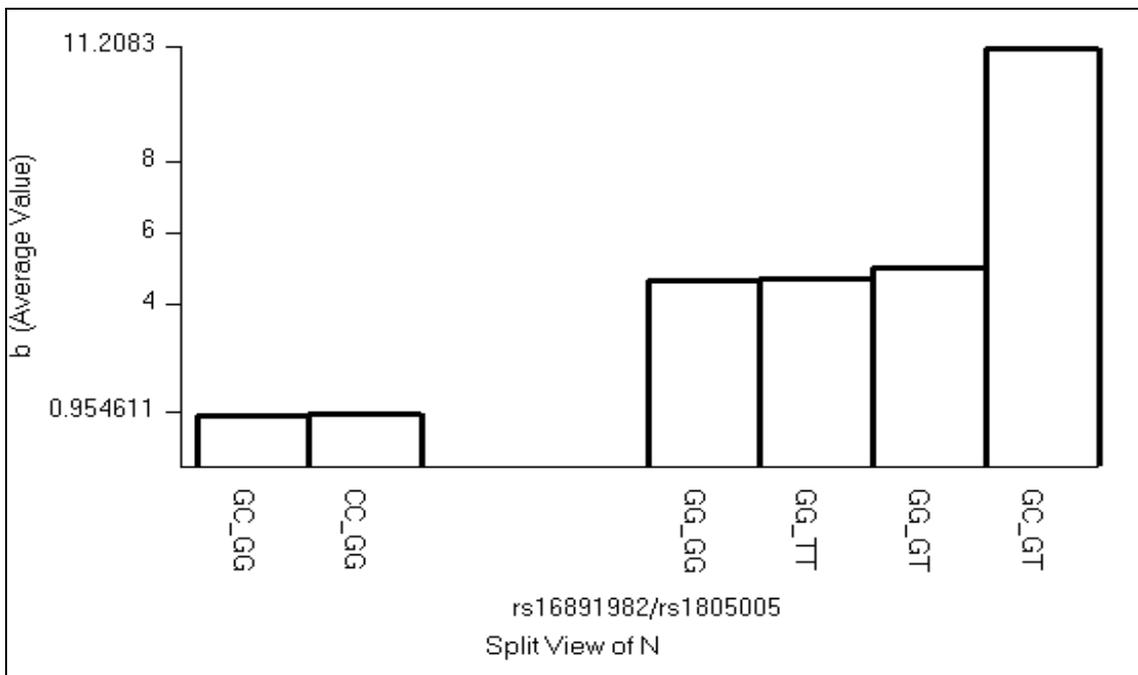


Figure 7.6 The associations between the rs16891982 and rs1805005 dihybrid and b* value

7.4 Discussion

7.4.1 Associations with Categorical Variables (Clusters)

In this analysis, five SNPs show various significant associations, in this population, with the hair colours as defined by the cluster analysis discussed in Chapter 3. These associations will be discussed; noting that D represents the minor allele and d represents the major allele found in this population. Since all SNPs in this study were non-synonymous mutations, the SNP will be referred to by the amino acid substitution listed in the Entrez SNP database [97].

As expected, a SNP from the MC1R gene (R151C) is associated with Red hair (Cluster 1) at the allelic level, as is a SNP from the TYR gene (R402Q). The high odds ratio of this SNP in MC1R (13.1) indicates the strength of this relationship. MC1R R151C is well known as a highly penetrant red-hair allele [14].

There are allelic associations with Fair/Blond hair in the genes DTNBP1 (H297Y) and MYO7A (C1628S). There are also dominant and recessive genotype associations with these two SNPs, as well as in MC1R (R151C), TYR (R402Q) and ATP7B (R832K).

When Cluster 1 (Red) is removed from the sample, the SNPs previously associated with Red hair are now associated with Fair/Blonde hair; however the odds ratios are much weaker (i.e. 2.3 compared to 13.1 for rs1805007). R832K in ATP7B is no longer associated with Cluster 2 when Cluster 1 is removed. The strongest associations (highest odds ratios) found within this sample were for H297Y in DNTBP1 in a dominant model (DD/Dd vs. dd) with an OR of 8.0 (when Reds were included in the sample) and R402Q in TYR in a recessive model with an OR of 3.8 (when Reds were excluded from the sample).

Associations were also found with Cluster 5 (Dark hair) in this sample. A SNP in the gene MATP (L374F) was significantly associated with dark hair in both the allelic and dominant model, which corresponds to previously published work on this gene [45]. Two SNPs in ATP7B (S406A and V456L) were also found to be associated in a

dominant model with Cluster 5 (Dark hair). The odds ratio for these two SNPs was less than one (0.49); therefore it is the major allele (d) that is associated with dark hair for these two polymorphisms and the 406S and 456V amino acid status. It should be noted that these are different SNPs than were associated with Blonde/Fair hair (that was R832K) and not just opposite alleles.

Of the six genes found to be significantly associated with hair colour clusters, several are familiar genes in the pigmentation pathway and are mentioned in Section 7.1.1. TYR codes for the rate-limiting melanogenesis protein, Tyrosinase [86], while MC1R (the melanocortin 1 receptor) is primarily involved with the switch between eumelanin and pheomelanin [4, 14, 86, 90]. MATP (membrane associated transport protein) is also referred to as SLC45A2, a solute carrier and a melanocyte differentiation antigen, thought to be involved in the mediation of melanin synthesis [45, 97].

Less familiar genes include DTNBP1, the gene for dystrobrevin-binding protein 1, which has been found to be associated with human skin pigmentation [109] and is thought to be involved in organelle biogenesis, including that of melanosomes [97]. The H297Y polymorphism in DTNBP1 was quite rare in this population, with a minor allele (A) frequency of 0.008 and this may contribute to the strong effect seen for this allele.

The gene ATP7B also appears to be associated with hair colour clusters. This gene codes for a Copper transporting ATPase [97, 110] and has been implicated in pigmentation of *Drosophila melanogaster* (fruit flies) [110]. The ATP7B protein is thought to control the exit of copper from the cells (Wilson disease in humans is caused by a defective ATP7B protein and results in toxic copper accumulation) [110]. Tyrosinase is a copper-dependant enzyme [111] and therefore; variations in cellular copper caused by polymorphisms in ATP7B may affect Tyrosinase activity.

The gene MYO7A codes for a Myosin VIIA protein, a mechanochemical protein that, in complex with Rab27a, is involved with the movement and distribution of melanosomes in the retinal pigmented epithelium [112] and may have similar effects in other pigmented tissues, such as hair.

7.4.2 Associations with Quantitative Variables (L^* , a^* and b^*)

Similar results to Section 7.4.1 were obtained when tests for association with the quantitative variables L^* , a^* and b^* were performed with various models – allelic, genotypic, dominant and recessive.

Polymorphisms in the genes TYR, MATP, DTNBP1, ATRN, MC1R, MLPH, MYO7A and MYO5A were found to be associated with L^* values and these results tend to correspond with genes that were associated with light or dark hair in Section 7.4.1. For example, polymorphisms in MATP that were associated with Dark Cluster 5 have a negative correlation to L^* , as would be expected. Seven new polymorphisms (not associated in the categorical analysis) were significantly associated with L^* in at least one of the models and these include two SNPs in the gene ATRN (I426T and R1152K), three SNPs in the gene MLPH (G172D, L153P and T289I), a different SNP in MYO7A (T16S) and one in MYO5A (R1246C).

The ATRN gene codes for the attractin protein; the mouse ortholog of this protein (*mahogany*) is involved in the melanocortin 1 receptor pathway and therefore may have an effect on the pheomelanin production [113].

The melanophilin protein is coded by MLPH and forms a motor complex with Rab27a and Myosin 5A (MYO5A) that has been found in mice to tether the melanosomes to the cytoskeleton for movement around the cell [94].

Tests for associations with a^* achieved very similar results to those with L^* with two notable exceptions. R151C in MC1R shows a much higher correlation ($R=0.43$) with a^* compared to other correlations in this sample set; indicating that it is a strong predictor of redness in hair, as expected. A second SNP in DTNBP1 (P272S) shows a small negative correlation with a^* .

The same polymorphisms continue to be associated with b^* , with the inclusion of H1207R in ATP7B and A420T in RABGGTA.

RABGGTA is a Rab geranylgeranyl transferase α sub-unit that is involved in the prenylation (adding lipids) of a Rab protein and is thought to anchor it to a membrane [114] and therefore is somewhat involved in the transfer of melanosomes.

7.4.3 Genetic Interactions

Two polymorphisms, one in the MATP gene (L374F) and one in MC1R (V60L) were shown to interact in their contribution to Fair/Blonde hair (they are associated with LL5 Cluster 2). It is possible that MATP;L374F may act as a suppressor of the fair hair in MC1R;V60L individuals, however many more samples would need to be analysed to elucidate the cause of any possible interactions. V60L is a different SNP in MC1R than is associated with red hair colour and is known as a low-penetrance red hair allele [14, 96].

7.4.4 Conclusions

It is clear that even in this small sample, there are significant associations with hair colour as defined by the cluster analysis in Chapter 3, particularly associations with Red, Fair/Blonde and Dark hair. There are also many significant correlations between genotypes at various SNPs and the quantitative measurements of L^* , a^* and b^* . As expected, one polymorphism or even one gene does not provide all of the information about hair colour and many genes and multiple polymorphisms within various genes seem to have an effect on hair colour. Most of these genes are familiar to those studying pigmentation; however, several are novel in the context of human hair pigmentation, though all have a logical connection to the pigmentation pathway. An expansion of the sample size will aid in further examination of these polymorphisms and may clarify and strengthen the associations found. An increase in the number of samples that were examined microscopically and chemically would allow for the further testing of genetic associations with these features, independently of their macroscopic colour values and clusters. With the current number of samples that have been microscopically and chemically analysed, no associations or trends were observed (data not shown). Additional SNPs should also be considered in the future. Several other

SNPs have recently been found by others to be associated with self-reported 'Blonde' hair [46], including additional SNPs in the genes SLC24A4, KITLG, TYR and OCA2 that have not been tested in this population or in another objectively measured one.

Another important next step in this work would be to conduct blind population assignment tests, to determine if likely genotype combinations can correctly assign individuals to a particular group or cluster. This approach would further validate results of the genetic associations and advance the practical applications of this type of work.

Chapter 8 Discussion and Conclusions

8.1 Summary

It was hypothesised in Chapter 1 that there are genetic variants specifically associated with an objectively defined Blonde phenotype. This project was conducted with an aim to first characterise and define the phenotype under study ('blonde hair') and then to evaluate DNA sequence variants that may be associated with this phenotype. These variants may be useful in predicting the hair colour of an unknown subject in forensic cases where DNA is found at a crime scene, but there is no suspect sample to compare it to.

The previous chapters have examined hair colour at several levels in accordance with the project aims. Macroscopic colour was measured with reflective spectrophotometry and digital image analysis, followed by a microscopic analysis of hair colour and pigmentation pattern. Next, a chemical characterisation was carried out, followed by analysis of genetic variations within genes thought to be involved in the pigmentation process.

Chapter 3 investigated the measurement of macroscopic colour by reflective spectrophotometry. This analysis showed that while hair colour can be perceived as a continuum, it can be divided into categories or clusters based on spectrophotometric measurements in the CIE $L^*a^*b^*$ colour space and these clusters can be well discriminated from each other. The importance of the objective measurement of individuals to be included in research studies is indicated by the discrepancies between self-reported hair colour and observer-reported hair colour and the poor separation of observer-reported colour groups compared to those defined by clustering analysis. The most favourable cluster grouping achieved in this analysis, for the purpose of this study, is called LL5 - when the log likelihood method is used to break individuals into five hair colour groupings, named Dark, Medium, Fair, Red and White.

Chapter 4 examined the use of digital image analysis to measure hair colour. This analysis showed that although measuring colour from digital images is convenient, the

colours measured and classified do not correspond very well to those colours measured by reflective spectrophotometry and would have limited use in studies requiring accurate measurement of hair colour, such as genetic association studies. They also do not provide an accurate description of hair colour for crime investigators to utilise as visual identifiers or in description announcements.

The microscopic colour of hair was examined in Chapter 5 and from this analysis, it can be seen that there is a strong relationship between macroscopic and microscopic L^* values that would be useful if one was trying to predict the colour of a head of hair (macroscopic) from a strand (microscopic). The relationship between hair lightness (L^*) and thickness is also clearly illustrated at both the macroscopic and microscopic levels.

The results of the chemical characterisation described in Chapter 6 show how the amount of eumelanin in hair (represented by its degradation product – PTCA) is related to the colour as measured in the CIE $L^*a^*b^*$ colour space by reflective spectrophotometry and microscopic analysis. As would be expected, lighter hair has less eumelanin; however, there is a stronger relationship between the amount of eumelanin present and the value of the b^* (yellow) axis than with the L^* or a^* axes. The amount of PTCA measured is more closely related to the microscopic colour than to the macroscopic colour, which may be due to the different methods of colour measurement (reflective spectrophotometry compared to a digital microscopic image). The PTCA results obtained for the LL5 clusters are also as expected. Clusters 5 (dark), 4 (Medium) and 2 (Fair) have progressively less of the eumelanin degradation product. The quantification of the pheomelanin degradation product (AHP) was unsuccessful in this sample in the time available for analysis.

Chapter 7 involves the genetic analysis with regards to both categorical variables - the LL5 clusters - and the quantitative variables – the L^* , a^* and b^* colour values. In the categorical analysis, five of 58 SNPs show various significant associations with Fair/Blonde hair (Cluster 2). These SNPs are in the genes MC1R (R151C), DTNBP1 (H297Y), TYR (R402Q), ATP7B (R832K) and MYO7A (C1628S). MC1R (R151C) and TYR (R402Q) were found to be associated with Red hair (Cluster 1) and associations were also found with Cluster 5 (Dark hair) in this sample. A SNP in the

gene MATP (L374F) was significantly associated with dark hair in both the allelic and dominant model, which corresponds to previously published work on this gene [45]. Two SNPs in ATP7B (S406A and V456L) were also found to be associated in a dominant model with Cluster 5 (Dark hair).

In the quantitative variable analysis, polymorphisms in the genes TYR, MATP, DTNBP1, ATRN, MC1R, MLPH, MYO7A and MYO5A were found to be associated with L* values and these results tend to correspond with genes that were associated with light or dark hair in the categorical analysis. Seven new polymorphisms (not associated in the categorical analysis) were significantly associated with L* in at least one of the models and these include two SNPS in the gene ATRN (I426T and R1152K), three SNPs in the gene MLPH (G172D, L153P and T289I), a different SNP in MYO7A (T16S) and one in MYO5A (R1246C).

Tests for associations with a* achieved very similar results to those with L* with two notable exceptions. R151C in MC1R shows a very high correlation ($R=0.43$) with a* compared to other correlations in this sample set; indicating that it is a strong predictor of redness in hair, as expected. A second SNP in DTNBP1 (P272S) shows a small negative correlation with a*. The same polymorphisms continue to be associated with b*, with the inclusion of H1207R in ATP7B and A420T in RABGGTA.

Two polymorphisms, one in the MATP gene (L374F) and one in MC1R (V60L) were shown to interact in their contribution to Fair/Blonde hair (they are associated with LL5 Cluster 2).

8.2 Discussion

It was hypothesised in Chapter 1 that there were genetic variants specifically associated with and objectively defined Blonde phenotype. To test this hypothesis, one must first define the phenotype under question and then search for any genetic associations. With this in mind, there were three primary objectives for this research project (listed in Chapter 1). The first objective was to examine the natural, adult hair colour variation among a population of European ancestry and to evaluate methods for both the

measurement and classification of hair colours – what is the phenotype and what are its limits?

The results of Chapters 3 and 4 fulfilled this objective very well. Natural, adult hair colour was studied extensively and the fair-to-dark hair colour spectrum has been characterised more thoroughly than previously reported in the literature [52, 59]. Blonde or Fair hair colour can be defined in several ways, by self or observer-reported colours or by various statistical clustering analyses using measurements such as reflective spectrophotometry or digital image analysis. This study found that objective measurements taken by reflective spectrophotometry and subject to statistical cluster analysis provided groupings that were objective, describable and useful as a phenotype due to the resemblance these groupings had to our subjective descriptions of hair colour. Even if the same clusters are not found in future samples (as is expected given changes in sample size and in various populations), the method can be used to achieve objectivity in classifying hair colours and is more likely to lead to an easier description of individuals and comparisons between research studies. If a large standardised database was to be constructed, particular groupings could be defined and then used by researchers to achieve consistency in their investigations. Such a database could also be used by forensic investigators and police to achieve consistent descriptions of hair colour.

There are a few other studies that have objectively measured hair colour [41, 52, 59] and having this data allows for potential comparisons between phenotypes and results of genetic studies. However; previous studies did not examine hair colour to the extent done here. As reported in Chapter 3 (Section 3.1.2), hair colour has been measured by reflective spectrophotometry; however these studies have used only the L* axis or the M index [59] or have concentrated on red hair [41, 52]. One study carefully measured hair colour (including red and blonde hair in particular) by reflective spectrophotometry, but used subjects of 12 and 14 years of age [41], running the risk of the subjects hair colour changing due to hormonal influences during puberty, such as the darkening of blonde hair [51].

Being able to measure hair colour from digital photographs may assist in efforts to construct or expand such a database and may also assist in various other areas with an

interest in hair colour. This may speculatively include fields such as surveillance, cosmetics or others. In this study, hair colour measured by digital image analysis did not correspond well to the reflective spectrophotometry standard and showed a high amount of variation, however; it may be possible for improvements in technology or software to change this outcome.

There are, of course, problems in the measurement of hair colour and in the future use of the data. While all volunteers in this study were healthy young adults with their natural hair colour, it is always possible that hair colours may be changed due to non-genetic influences such as sunlight, hormonal changes and other epigenetic factors and the use of hair products [16, 17, 51]. While every effort was made to reduce the influence of these processes for this early-stage research, they will have effects on hair colour in practice and further research into each of these factors may assist in the characterisation of hair colours, which will eventually assist in any use of predictive forensic DNA tests or the use of hair colour data in traditional hair examinations. In addition, not all DNA samples tested will belong to young healthy adults and so further research into hair colour changes due to age or medical conditions may also be worthwhile pursuing.

The second objective was to characterise the microscopic and chemical features of hair colour in this sample and how they relate to the macroscopic colour – what are the properties of this phenotype? The microscopic analysis undertaken for this project (Chapter 5) provided some very interesting, new results. While microscopic hair colour has been examined to a small extent, in an effort to assist with traditional hair examinations and identification of hairs [24, 27], there seem to be no previous studies relating the microscopic hair colour to the macroscopic, a deficit in the literature that this research will help fill. Traditional forensic hair examiners often offer an opinion to the hair colour of an individual based on a microscopic analysis of a strand of hair [22, 115]. Having statistical data to support such claims will assist in this type of examination. It may also help link the various levels of analysis – genetic, chemical, microscopic and macroscopic – providing information to the scientist trying to predict hair colour based on any of the above data (particularly genetic). Due to the significant results achieved in this experiment, it would be beneficial to continue the analysis on an expanded sample size. The examination of more individuals and more hairs from each

individual would assist in the further analysis of the relationship between microscopic and macroscopic hair colour.

The chemical analysis in Chapter 6 supported previous research by finding similar eumelanin levels in various hair colour categories, as discussed in Section 6.4 [20, 82], but having the macroscopic and microscopic hair colour measurements allowed for a more specific relationship to be established between the amount of PTCA recovered (the eumelanin degradation product) and the colour of the hair. Chemically characterising more of the individuals in this sample or on a different, expanded sample will further elucidate relationships between pigment composition and colour. It is widely accepted that less eumelanin pigment leads to a lighter hair colour and this experiment does not show anything unexpected. It does, however, provide data for this particular sample and it provides chemical characterisation of the reflective spectrophotometry clusters. This may be one way for various researchers to compare their work, if colour measurements were not taken. If enough samples could be collected for chemical analysis, it would be worthwhile conducting genetic association tests directly on the amount of pigment present in the hair samples. Unfortunately, the 55 individuals in this sample with chemical data do not provide enough power for meaningful genetic analysis but should be reconsidered in the future.

It would also be beneficial to obtain pheomelanin characterisation data for this sample. Even though eumelanin is the primary pigment of interest when studying the light-to-dark continuum of hair colour, the ratio between the eumelanin and pheomelanin pigments is relevant to the study of hair colour genetics. In the time available for this analysis in the current study, the analysis of the pheomelanin pigments could not be made to work. Further efforts would eventually change this outcome and the data could be incorporated into the analysis.

The third and final objective was to compare genetic results with an objectively measured and defined hair colour phenotype in the search for genetic associations with 'blonde' hair colour. This was a small, but well-characterised sample and as reported in Chapter 7, several significant genetic associations were found with Fair/Blonde hair as defined in this cluster analysis. These results support other genetic studies that have been released recently [45, 46] by finding associations in some of the same genes (such

as MC1R and TYR), with the small effects expected in the genetic analysis of hair colour [9]. This study also offered some new associations that would be worth examining in the future on a larger sample. The genes that showed significant associations in this study were in various areas of the pigmentation pathway, including the manufacture of melanin (melanogenesis) as expected, but many were involved in other areas of pigmentation, such as the formation of melanosomes and the movement of melanosomes around the cell. These genes seem to be more involved in the light-to-dark differences in pigmentation within individuals of European ancestry than those genes typically involved in population differences (melanogenesis [4, 34]) and red hair (pheomelanin switch [4, 34]). Looking at other genes involved in melanosome biogenesis and movement and examining the effect of causal variations in these genes would be an important part of future research into genetics and cellular biology of pigmentation. As well as examining additional candidate genes, looking for associations with clusters or quantitative colour measurements using whole genome scans may provide additional information in the search for hair colour genes.

There is a great deal of work left before the prediction of hair colour from genetic information is remotely possible, however this study will provide future researchers with associations that can be expected in an objectively measured population. Large, objectively measured samples are difficult to come by. If developments in technology allow for the easier measurements of hair colour (by any method), this will become more likely and genetic analysis may become more accurate. Without objective measurements, the prediction of physical features for forensic prediction is still possible; however it will be less accurate and may be open to more challenges from investigators and the justice system.

The results of this study address some of the issues in human hair colour genetics that were raised in Section 2.3.3. This study provides additional evidence that the genes involved with hair colour are somewhat different to those most strongly associated with eye colour and that there is a larger number of gene variants with lower penetrance involved in hair colour. 19 SNPs were found to have some association with hair colour in this study and most of these SNPs had fairly small effects (odds ratios of 1 to 4 or correlations between 0.1 and 0.3), a trend which was also seen in the deCODE study mentioned previously [46]. As well, OCA2 (previously shown to be the main eye

colour gene [116]) does not seem to have a strong enough effect on hair colour to be seen in this sample, if any, supporting the idea that different genes are involved.

There is also some indication that epistatic genetic interaction may contribute to differences in hair colour, such as the interaction found between MATP and MC1R in this study and a previous interaction between MC1R and a SNP in the 3' untranslated region of the ASIP gene [44]. Although one interaction was detected in this sample, due to generally low gene effects, the study population would likely have to be much larger to find any more such interactions.

One of the issues in finding genetic associations with hair colour that has been mentioned in the literature [9] is that the phenotyping of hair may not be 'specific' enough to find consistent and specific associations. While this still may be true, to some degree, this project has provided a great deal of information on objectively defining hair colour phenotypes and offers methods and data for the future description of hair colours in various types of research studies.

8.3 Future Considerations

As a pilot study, the presented research has raised some interesting ideas that would benefit from further examination. The most important consideration for future research would be to expand the sample size. All results obtained in this study are statistically significant; however, an increase in the sample size can only improve the power of analysis, particularly for the genetic analysis. An increase in the number of individuals examined by microscopic and chemical analysis would also benefit further studies by more strongly illustrating existing relationships. Increasing the number of individuals with objectively measured hair colour would assist in the development of a 'phenotypic database' – a database with images of all possible hair colours. Sample sizes consisting of thousands of individuals (such as in the deCODE genetics study discussed in Chapters 2 and 7) will eventually be necessary for both hair colour research and genetic association tests and the challenge of objectively measuring the hair colour of that many individuals will take a lot of work. The use of more modern reflective spectrophotometers (that can output results directly to a notebook computer) would

assist in this effort and do exist, but were not available for this study. The development of digital photographic equipment and colour measurement software specific to measuring hair colour may also be helpful.

Another area of research that would be useful and interesting is the further study of the cellular biology of hair pigmentation. One study by Commo *et al.* (2004) examined the distributions of different proteins in African, Asian and Caucasian hair bulbs and as a part of their study, tagged the pMel-17 melanosome identifier protein [50]. Since some of the genetic variations associated with blonde hair are in genes that control the biogenesis and movement of melanosomes, it may be worth using a similar methodology to examine the distribution of these organelles in hair bulbs within a population of European ancestry across various hair colours to determine if the number or distribution of melanosomes is correlated with hair colour. This would increase the understanding of the underlying pigmentation processes and would further elucidate the specific genetic contributions to hair colour.

Further examination of the exact composition of hair melanin pigments would be another area of research that would assist in the characterisation of hair for forensic or genetic purposes. While perhaps unlikely, improved availability of and information on melanin degradation product standards (PTCA, PDCA and AHP) would assist in method development and allow for more thorough analysis.

Another logical progression from this study would be the genotyping of several SNPs that have been associated with fair or blonde hair colour in the literature published since this set of 58 SNPs was chosen. These include SNPs in the genes SLC24A4, KITLG, TYR, OCA2 and MC1R [46] and possibly in other genes implicated. It may also be worth looking at more SNPs in genes found to have significant associations in this study, there may be a haplotype effect to be considered or these SNPs may be in linkage disequilibrium with the causative mutation. As well as finding significant associations, discovering how these causative mutations affect proteins produced and how they may interact with other genes will help researchers understand hair colour and will lead to more accurate predictions.

8.4 Conclusion

In conclusion, this study successfully met its aims of objectively classifying and characterising blonde (or fair) hair colour and finding genetic associations with blonde hair colour. Some of the data supports previous studies in the area of hair colour characterisation and genetics and a great deal of it is new and can be useful to researchers in forensic science and in other areas of pigmentation research. There are several matters that may be expanded on in the future and it is hoped that this study may be a resource to those conducting hair pigmentation research in anthropology, genetics and in forensic science.

Colours are the smiles of nature.

--Leigh Hunt (1784-1859)

Chapter 9 References

1. Sargent, W., *The enjoyment and use of color*. 1923, New York, NY: Charles Scribner's Sons.
2. Westerhof, W., *The discovery of the human melanocyte*. *Pigment Cell Research*, 2006. **19**: p. 183-193.
3. Sturm, R.A., N.F. Box, and M. Ramsay, *Human pigmentation genetics: The difference is only skin deep*. *BioEssays*, 1998. **20**: p. 712-721.
4. Rees, J.L., *Genetics of hair and skin color*. *Annual Reviews Genetics*, 2003. **37**: p. 67-90.
5. Menasche, G., C.H. Ho, O. Sanal, J. Feldman, I. Tezcan, F. Ersoy, A. Houdusse, A. Fischer, and G. de Saint Basile, *Griscelli syndrome restricted to hypopigmentation results from a melanophilin defect (gs3) or a myo5a f-exon deletion (gs1)*. *The Journal of Clinical Investigation*, 2003. **112**: p. 450-456.
6. Slominski, A., J. Wortsman, P.M. Plonka, K.U. Schallreuter, R. Paus, and D.J. Tobin, *Hair follicle pigmentation*. *Journal of Investigative Dermatology*, 2004. **124**: p. 13-21.
7. Duffy, D.L., N.F. Box, W. Chen, J.S. Palmer, G.W. Montgomery, M.R. James, N.K. Hayward, N.G. Martin, and R.A. Sturm, *Interactive effects of mc1r and oca2 on melanoma risk phenotypes*. *Human Molecular Genetics*, 2004. **13**(4): p. 477-461.
8. Parra, E.J., R.A. Kittles, and M.D. Shriver, *Implications of correlations between skin color and genetic ancestry for biomedical research*. *Nature Genetics*, 2004. **36**(11): p. s54-s60.
9. Frudakis, T.N., *Molecular photofitting*. 2008, Burlington, MA: Elsevier Inc.
10. Frudakis, T.N., M. Thomas, Z. Gaskin, K. Venkateswarlu, K. Suresh Chandra, S. Ginjupalli, S. Gunturi, S. Natrajan, V.K. Ponnuswamy, and K.N. Ponnuswamy, *Sequences associated with human iris pigmentation*. *Genetics*, 2003. **165**: p. 2071-2083.
11. Frudakis, T.N., K. Venkateswarlu, M.J. Thomas, Z. Gaskin, S. Ginjupalli, S. Gunturi, V.K. Ponnuswamy, S. Natrajan, and K.N. Ponnuswamy, *A classifier for the snp-based inference of ancestry*. *Journal of Forensic Science*, 2003. **48**(4): p. 1-12.
12. Frudakis, T.N., M. Thomas, and Z. Gaskin, *Snp based ancestry informative markers for the inference of biogeographical ancestry (i.E., "race") and the estimation of admixture: Application for the louisiana serial killer case, in American Academy of Forensic Science*. 2004: Dallas, TX. United States.

13. Baidur-Hudson, S. and R. van Oorschot, *Personal communication*. 2005.
14. Ha, T., L. Naysmith, K. Waterson, C. Oh, R. Weller, and J.L. Rees, *Defining the quantitative contribution of the melanocortin 1 receptor (mc1r) to variation in pigmentary phenotype*. *Annals of the New York Academy of Science*, 2003. **994**: p. 339-347.
15. Ancans, J., N. Flanagan, M.J. Hoogduijn, and A.J. Thody, *P-locus is a target for the melanogenic effects of mc1r signalling: A possible control point for facultative pigmentation*. *Annals of the New York Academy of Science*, 2003. **994**: p. 373-377.
16. Takahashi, T. and K. Nakamura, *A study of the photolightening mechanism of blond hair with visible and ultraviolet light*. *Journal of Cosmetic Science*, 2004. **55**: p. 291-305.
17. Takahashi, T. and K. Nakamura, *A study of the photolightening mechanism of red hair with visible and ultraviolet light: Comparison with blond hair*. *Journal of Cosmetic Science*, 2005. **56**: p. 47-56.
18. Grimes, E.A., P.J. Noake, L. Dixon, and A. Urquhart, *Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype*. *Forensic Science International*, 2001. **122**: p. 124-129.
19. Nogueira, A.C.S. and I. Joekes, *Hair color changes and protein damage caused by ultraviolet radiation*. *Journal of Photochemistry and Photobiology B: Biology*, 2004. **74**: p. 109-117.
20. Borges, C.R., J.C. Roberts, D.G. Wilkins, and D.E. Rollins, *Relationship of melanin degradation products to actual melanin content: Application to human hair*. *Analytical Biochemistry*, 2001. **290**: p. 116-125.
21. Saferstein, R., *Criminalistics*. 7th ed. 2001, Upper Saddle River, New Jersey: Prentice Hall.
22. Robertson, J., *Forensic examination of human hair*. Taylor and Francis forensic science series, ed. J. Robertson. 1999, London: Taylor and Francis.
23. FBI, *Forensic human hair examination guidelines*. 2005, Federal Bureau of Investigation: Forensic Science Communications: Washington, DC.
24. Bednarek, J., *An attempt to establish objective criteria for morphological examinations of hairs using the image analysis system*. *Problems of Forensic Science*, 2004. **LVI**: p. 65-77.
25. Cicerone, R.J. and C.M. Vest. *Strengthening forensic science in the united states: A path forward*. 2009 [cited 2010 March 7]; Available from: http://books.nap.edu/openbook.php?record_id=12589&page=183.
26. Cooper, A., *Personal communication*. 2008, University of Adelaide.

27. Brooks, E., *An appraisal of the use of numerical features in the forensic examination of hair (msc thesis)*, in *School of Health Sciences*. 2007, University of Canberra. <http://erl.canberra.edu.au/public/adt-AUC20080624.144159/index.html>
28. Butler, J.M., *Forensic DNA typing: Biology, technology, and genetics of str markers* 2005, Burlington, MA USA: Elsevier.
29. Jobling, M.A. and P. Gill, *Encoded evidence: DNA in forensic analysis*. *Nature Reviews Genetics*, 2004. **5**: p. 739-751.
30. Ribaux, O., A. Girod, S.J. Walsh, P. Margot, S. Mizrahi, and V. Clivaz, *Forensic intelligence and crime analysis* *Law, Probability and Risk*, 2003. **2**(1): p. 47-60.
31. Tully, G., *Genotype versus phenotype: Human pigmentation*. *Forensic Science International: Genetics*, 2007. **1**: p. 105-110.
32. Daniel, R., J.J. Sanchez, N.T. Nassif, A. Hernandez, and S.J. Walsh, *Snps associated with physical traits: A valuable tool for the inference of biogeographical ancestry*. *Forensic Science International: Genetics Supplement Series*, 2008. **1**(1): p. 538-540.
33. Daniel, R., J.J. Sanchez, N.T. Nassif, A. Hernandez, and S.J. Walsh, *Partial forensic validation of a 16plex snp assay for the inference of biogeographical ancestry*. *Forensic Science International: Genetics Supplement Series*, 2009. **2**(1): p. 477-478.
34. Parra, E.J., *Human pigmentation variation: Evolution, genetic basis, and implications for public health*. *Yearbook of Physical Anthropology*, 2007. **50**: p. 85-105.
35. Rebbeck, T.R., P.A. Kanetsky, A.H. Walker, R. Holmes, A.C. Helpert, L.M. Schuchter, D.E. Elder, and D. Guerry, *P gene as an inherited biomarker of human eye color*. *Cancer Epidemiology, Biomarkers and Prevention*, 2002. **11**: p. 782-784.
36. Valverde, P., E. Healy, I.J. Jackson, J.L. Rees, and A.J. Thody, *Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans*. *Nature Genetics*, 1995. **11**: p. 328-330.
37. Beaumont, K.A., R.A. Newton, D.J. Smit, J.H. Leonard, J.L. Stow, and R.A. Sturm, *Altered cell surface expression of human mclr variant receptor alleles associated with red hair and skin cancer risk*. *Human Molecular Genetics*, 2005. **14**(15): p. 2145-2154.
38. Box, N.F., J.R. Wyeth, L.E. O'Gorman, N.G. Martin, and R.A. Sturm, *Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair*. *Human Molecular Genetics*, 1997. **6**(11): p. 1891-1897.

39. Harding, R.M., E. Healy, A.J. Ray, N.S. Ellis, N. Flanagan, C. Todd, C. Dixon, A. Sajantila, I.J. Jackson, M.A. Birch-Machin, and J.L. Rees, *Evidence for variable selective pressures at mc1r*. American Journal of Human Genetics, 2000. **66**: p. 1351-1361.
40. Voisey, J., N.F. Box, and A. Van Daal, *A polymorphism study of the human agouti gene and its association with mc1r*. Pigment Cell Research, 2001. **14**: p. 264-267.
41. Shekar, S.N., D.L. Duffy, T.N. Frudakis, G.W. Montgomery, M.R. James, R.A. Sturm, and N.G. Martin, *Spectrophotometric methods for quantifying pigmentation in human hair - influence of mc1r genotype and environment*. Photochemistry and Photobiology, 2008. **84**: p. 719-726.
42. Healy, E., S.A. Jordan, P. Budd, S., R. Suffolk, J.L. Rees, and I.J. Jackson, *Functional variation of mc1r alleles from red-haired individuals*. Human Molecular Genetics, 2001. **10**(21): p. 2397-2402.
43. Bonilla, C., L.-A. Boxill, S.A. Mc Donald, T. Williams, N. Sylvester, E.J. Parra, S. Dios, H.L. Norton, M.D. Shriver, and R.A. Kittles, *The 8818g allele of the agouti signaling protein (asip) gene is ancestral and is associated with darker skin color in african americans*. Human Genetics, 2005. **116**: p. 402-406.
44. Kanetsky, P.A., J. Swoyer, S. Panossian, R. Holmes, D. Guerry, and T.R. Rebbeck, *A polymorphism in the agouti signaling protein gene is associated with human pigmentation*. American Journal of Human Genetics, 2002. **70**: p. 770-775.
45. Graf, J., R. Hodgson, and A. van Daal, *Single nucleotide polymorphisms in the matp gene are associated with normal human pigmentation variation*. Human Mutation, 2005. **25**: p. 278-284.
46. Sulem, P., D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Palsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, L.A. Kiemenev, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, and K. Stefansson, *Genetic determinants of hair, eye and skin pigmentation in europeans*. Nature Genetics, 2007. **39**(12): p. 1443-1452.
47. Ito, S. and K. Wakamatsu, *Quantitative analysis of eumelanin and pheomelanin in humans, mice, and other animals: A comparative review*. Pigment Cell Research, 2003. **16**: p. 523-531.
48. Ozeki, H., S. Ito, K. Wakamatsu, and T. Hirobe, *Chemical characterisation of hair melanins in various coat-colour mutants of mice*. Journal of Investigative Dermatology, 1995. **105**: p. 361-366.

49. Norton, H.L., J.S. Friedlaender, A. Merriwether, G. Koki, C.S. Mgone, and M.D. Shriver, *Skin and hair pigmentation variation in island melanesia*. American Journal of Physical Anthropology, 2006. **130**(2): p. 254-268.
50. Commo, S., O. Gaillard, S. Thibaut, and B.A. Bernard, *Absence of trp-2 in melanogenic melanocytes of human hair*. Pigment Cell Research, 2004. **17**: p. 488-497.
51. Slominski, A., D.J. Tobin, S. Shibahara, and J. Wortsman, *Melanin pigmentation in mammalian skin and its hormonal regulation*. Physiological Reviews, 2004. **84**: p. 1155-1228.
52. Naysmith, L., K. Waterson, T. Ha, N. Flanagan, Y. Bisset, A. Ray, K. Wakamatsu, S. Ito, and J.L. Rees, *Quantitative measures of the effect of the melanocortin 1 receptor on human pigmentary status*. Journal of Investigative Dermatology, 2004. **122**: p. 423-428.
53. Fujimoto, A., R. Kimura, J. Ohashi, K. Omi, R. Yuliwulandari, L. Batubara, M.S. Mustofa, U. Samakkarn, W. Settheetham-Ishida, T. Ishida, Y. Morishita, T. Furusawa, M. Nakazawa, R. Ohtsuka, and K. Tokunaga, *A scan for genetic determinants of human hair morphology: Edar is associated with asian hair thickness*. Human Molecular Genetics, 2008. **17**(6): p. 835-843.
54. Dictionary.com. Unabridged (v 1.1). [cited; Available from: <http://dictionary.reference.com/browse/color> (accessed: May 28, 2008).
55. Hunt, R.W.G., *Measuring colour*. 3rd ed. 1998, Kingston-upon-Thames, England: Fountain Press.
56. Hurvich, L.M. and D. Jameson, *An opponent-process theory of colour vision*. Psychological Review, 1957. **64**(6): p. 384-404.
57. TASI. *Colour theory: Understanding and modelling colour*. December 5, 2007 [cited; Available from: <http://www.tasi.ac.uk/advice/creating/pdf/colour.pdf>.
58. Ford, A. and A. Roberts. *Colour space conversions*. [cited February 1, 2007]; Available from: www.poynton.com/PDFs/coloureq.pdf.
59. Shriver, M.D. and E.J. Parra, *Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry*. American Journal of Physical Anthropology, 2000. **112**: p. 17-27.
60. Napier, S., *Personal communication*. 2007, Biolab Group Australia.
61. Cecil, *Principles and applications of reflectance spectrophotometry*, in *LabPlus International*. 2006, Cecil Instruments.

62. Jacobson, R.E., S.F. Ray, G.G. Attridge, and N.R. Axford, *The manual of photography: Photographic and digital imaging*. 2000, San Diego, USA: Focal Press: Elsevier Science Inc.
63. Herbin, M., A. Venot, J.Y. Devaux, and P. C., *Color quantitation through image processing in dermatology*. IEEE Transactions on Medical Imaging, 1990. **9**(3): p. 262-269.
64. Shen, Z., *Colour differentiation in digital images (msc thesis)*, in *School of Computer Science and Mathematics*. 2003, Victoria University. p167.
65. Levin, N., E. Ben-Dor, and A. Singer, *A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments*. International Journal of Remote Sensing, 2005. **26**(24): p. 5475 - 5492.
66. Kumara, B., S. Jinap, Y.B. Che Man, and M.S.A. Yusoff, *Note: Comparison of colour techniques to measure chocolate fat bloom*. Food Science and Technology International, 2003. **9**(4): p. 295-299.
67. Takiwaki, H., H. Miyamoto, and K. Ahsan, *A simple method to estimate $cie-l^*a^*b^*$ values of the skin from its videomicroscopic image*. Skin Research and Technology, 1997. **3**: p. 42-44.
68. Pierard, G.E., *Eemco guidance for the assessment of skin colour*. Journal of the European Academy of Dermatology and Venereology, 1998. **10**: p. 1-11.
69. Comber, B., *Personal communication*. 2007, Australian Federal Police.
70. Rasband, W.S., *Imagej*. 1997-2006, US National Institutes of Health, Bethesda, Maryland, USA, <http://rsb.info.nih.gov/ij/>.
71. McLave, J.T. and T. Sincich, *Statistics*. 9th ed. 2003, Upper Saddle River, NJ: Prentice-Hall Inc.
72. Vaughn, M., R. van Oorschot, and S. Baindur-Hudson, *Hair color measurement and variation*. American Journal of Physical Anthropology, 2008. **137**(1): p. 91-96.
73. Harding, H. and G. Rogers, *Forensic examination of human hair*. Taylor and francis forensic science series, ed. J. Robertson. 1999, London: Taylor and Francis.
74. Tobin, D.J., *Biochemistry of human skin—our brain on the outside*. Chemical Society Reviews, 2006. **35**: p. 52-67.
75. Van Den Bossche, K., J.-M. Naeyaert, and J. Lambert, *The quest for the mechanism of melanin transfer*. Traffic, 2006. **7**: p. 769-778.
76. Wakamatsu, K. and S. Ito, *Advanced chemical methods in melanin determination*. Pigment Cell Research, 2004. **15**: p. 174-183.

77. Ito, S. and K. Fujita, *Microanalysis of eumelanin and pheomelanin in hair and melanomas by chemical degradation and liquid chromatography*. Analytical Biochemistry, 1985. **144**: p. 527-536.
78. Napolitano, A., A. Pezzella, M.R. Vincensi, and G. Prota, *Oxidative degradation of melanins to pyrrole acids: A model study*. Tetrahedron, 1995. **51**(20): p. 5913-5920.
79. Gautam, L., K.S. Scott, T.F. Emmett, and M.D. Cole, *Characterisation of natural and synthetic melanins*. Global Forensic Science Today, 2008. **6**: p. 27-43.
80. Wakamatsu, K., S. Ito, and J.L. Rees, *The usefulness of 4-amino-3-hydroxyphenylalanine as a specific marker of pheomelanin*. Pigment Cell Research, 2002. **15**: p. 225-232.
81. Szekely-Klepser, G., K. Wade, D. Woolson, R. Brown, S. Fountain, and E. Kindt, *A validated lc/ms/ms method for the quantification of pyrrole-2,3,5-carboxylic acid (ptca), a eumelanin specific biomarker, in human skin punch biopsies*. Journal of Chromatography B, 2005. **826**: p. 31-40.
82. Napolitano, A., M.R. Vincensi, P.D. Donato, G. Monfrecola, and G. Prota, *Microanalysis of melanins in mammalian hair by alkaline hydrogen peroxide degradation: Identification of a new structural marker of pheomelanins*. Journal of Investigative Dermatology, 2000. **114**: p. 1141-1147.
83. Wakamatsu, K., A. Takasaki, B. Kagedal, T. Kageshita, and S. Ito, *Determination of eumelanin in human urine*. Pigment Cell Research, 2006. **19**: p. 163-169.
84. Nezirevic, D., K. Arstrand, and B. Kagedal, *Hydrophilic interaction liquid chromatographic analysis of aminohydroxyphenylalanines from melanin pigments*. Journal of Chromatography A, 2007. **1163**: p. 70-79.
85. Yang, Q., X.-L. Zhang, M. Ma, K.-J. Huang, J.-X. Zhang, W.-Z. Ni, C.-X. Fang, and C.-Y. Zheng, *New high-performance liquid chromatographic method for sensitive determination of pheomelanin in biological materials by precolumn fluorescence derivatization with naphthalene-2,3-dicarboxaldehyde* Journal of Chromatography A, 2007. **1146**(1): p. 23-21.
86. Sturm, R.A., R.D. Teasdale, and N.F. Box, *Human pigmentation genes: Identification, structure and consequences of polymorphic variation*. Gene, 2001. **277**: p. 49-62.
87. Box, N.F., J.R. Wyeth, C.J. Mayne, L.E. O'Gorman, N.G. Martin, and R.A. Sturm, *Complete sequence and polymorphism study of the human tyrp1 gene encoding tyrosinase-related protein 1*. Mammalian Genome, 1998. **9**: p. 50-53.

88. Smith, D.R., D.T. Spaulding, H.M. Glenn, and B.B. Fuller, *The relationship between na⁺/h⁺ exchanger expression and tryosinase activity in human melanocytes*. Experimental Cell Research, 2004. **298**: p. 521-534.
89. Brilliant, M.H., *The mouse p (pink-eyed dilution) and human p genes, oculocutaneous albinism type 2 (oca2), and melanosomal ph*. Pigment Cell Research, 2001. **14**: p. 86-93.
90. John, P.R., K. Makova, W.-H. Li, T. Jenkins, and M. Ramsay, *DNA polymorphism and selection at the melanocortin-1 receptor gene in normally pigmented southern african individuals*. Annals of the New York Academy of Science, 2003. **994**: p. 299-306.
91. Kauser, S., A.J. Thody, K.U. Schallreuter, C.L. Gummer, and D.J. Tobin, *A fully functional proopiomelanocortin/melanocortin-1 receptor system regulates the differentiation of human scalp hair follicle melanocytes*. Endocrinology, 2005. **146**(2): p. 532-543.
92. Oetting, W. and D.C. Bennett. *International albinism centre*. 2006 [cited November 22, 2006]; Available from: <http://albinismdb.med.umn.edu/>.
93. Oetting, W.S., S. Savage Garret, M.J. Brott, and R.A. King, *P gene mutations associated with oculocutaneous albinism type ii (oca2)*. Human Mutation, 2005.
94. Strom, M., A.N. Hume, A.K. Tarafder, E. Barkagianni, and M.C. Seabra, *A family of rab27-binding proteins*. The Journal of Biological Chemistry, 2002. **277**(28): p. 25423-25430.
95. Lamason, R.L., M.-A.P.K. Mohideen, J.R. Mest, A.C. Wong, H.L. Norton, M.C. Aros, M.J. Jurynec, X. Mao, V.R. Humphreville, J.E. Humbert, S. Sinha, J.L. Moore, P. Jagadeeswaran, W. Zhao, G. Ning, I. Makalowska, P.M. McKeige, D. O'Donnell, R.A. Kittles, E.J. Parra, N.J. Mangini, D.J. Grunwald, M.D. Shriver, V.A. Canfield, and K.C. Cheng, *Slc24a5, a putative cation exchanger, affects pigmentation in zebrafish and humans*. Science, 2005. **310**: p. 1782-1786.
96. Sturm, R.A., D.L. Duffy, N.F. Box, R.A. Newton, A.G. Shepherd, W. Chen, L.H. Marks, J.H. Leonard, and N.G. Martin, *Genetic association and cellular function of mc1r variant alleles in human pigmentation*. Annals of the New York Academy of Science, 2003. **994**: p. 348-358.
97. Database, E.S., National Center for Biotechnology Information.
98. van Daal, A., *Personal communication*. 2007, Bond University.
99. Budowle, B. and A. van Daal, *Forensically relevant snp classes*. BioTechniques, 2008. **44**(5): p. 603-610.
100. Qiagen. *Qiagen proteinase k*. 2010 [cited 2010 March 21]; Available from: <http://www1.qiagen.com/products/accessories/accessoryenzymes/qiagenproteina sek.aspx?ShowInfo=1>.

101. Koukoulas, I., F.E. O'Toole, P. Stringer, and R.A.H. van Oorschot, *Quantifiler™ observations of relevance to forensic casework*. Journal of Forensic Sciences, 2008. **53**(1): p. 135-141.
102. *Quantification of human genomic DNA using the quantifiler™ human DNA quantification kit and the abi prism® 7500 sds instrument*, VPFSC Biology Division, DNA Science Branch Method [MBIO] 31.
103. Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, *Current protocols in molecular biology*. Vol. 2. 2002, Hoboken, NJ Wiley & Sons Inc.
104. Sabatti, C., S. Service, and N. Freimer, *False discovery rate in linkage and association genome screens for complex disorders* Genetics, 2003. **164**: p. 823-833.
105. Parker, R., *Personal communication*. 2008, Golden Helix Inc.
106. *Entrez snp*. January 13, 2009 [cited; Available from: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp>].
107. *International hapmap project*. January 13, 2009 [cited; Available from: <http://www.hapmap.org/cgi-perl/gbrowse/>].
108. Healy, E., S.A. Jordan, P.S. Budd, R. Suffolk, J.L. Rees, and I.J. Jackson, *Functional variation of mc1r alleles from red-haired individuals*. Human Molecular Genetics, 2001. **10**(21): p. 2397-2402.
109. Voight, B.F., S. Kudravalli, X. Wen, and J.K. Pritchard, *A map of recent positive selection in the human genome*. PLoS Biology, 2006. **4**(3): p. e72.
110. Zhou, H., K.M. Cadigan, and D.J. Thiele, *A copper-regulated transporter required for copper acquisition, pigmentation, and specific stages of development in drosophila melanogaster*. The Journal of Biological Chemistry, 2003. **278**(48): p. 48210-48218.
111. Lerner, A.B., T.B. Fitzpatrick, E. Calkins, and W.H. Summerson, *Mammalian tyrosinase: The relationship of copper to enzymatic activity*. The Journal of Biological Chemistry, 1950. **187**(2): p. 793-802.
112. Gibbs, D., S.M. Azarian, C. Lillo, J. Kitamoto, A.E. Klomp, K.P. Steel, R.T. Libby, and D.S. Williams, *Role of myosin viia and rab27a in the motility and localization of rpe melanosomes*. Journal of Cell Science, 2004. **117**(266473-6483).
113. He, L., T.M. Gunn, D.M. Bouley, X.-Y. Lu, S.J. Watson, S.F. Schlossman, J.S. Duke-Cohan, and G.S. Barsh, *A biochemical function for attractin in agouti-induced pigmentation and obesity*. Nature Genetics, 2001. **27**: p. 40-47.

114. Maurer-Stroh, S. and F. Eisenhaber, *Refinement and prediction of protein prenylation motifs*. *Genome Biology*, 2005. **6**: p. R55.1 - R55.15.
115. SWGMAT, *Forensic human hair examination guidelines*. 2005, Forensic Science Communications: Washington, DC.
116. Sturm, R.A. and T.N. Frudakis, *Eye colour: Portals into pigmentation genes and ancestry*. *Trends in Genetics*, 2004. **20**(8): p. 327-332.

APPENDIX 1 – Raw Data

Individual	Reflective Spectrophotometry			Digital Image Analysis			Microscopic Analysis			Observer Reported Colour ¹	Hair Thickness (μm)	Chemical Analysis PTCA (ng/mg hair)
	L*	a*	b*	L*	a*	b*	L*	a*	b*			
HC01 ^z	34.00	3.48	-1.19	53.49	-2.53	8.39	9.39	8.02	6.05	NE	14.44	296.07
HC02	53.21	4.91	10.25	81.28	1.25	14.44	73.54	0.62	6.92	1	8.78	26.19
HC03 ^z	34.91	3.60	-0.41	56.34	-0.26	10.31	13.84	10.22	7.99	NE	19.22	200.99
HC04	30.78	1.83	1.57	56.5	0.51	16.17	46.28	8.44	15.20	3	15.44	104.86
HC05	46.55	4.11	10.85	77.95	3.55	26.03	61.87	4.79	12.55	1	11.56	72.73
HC06	26.01	0.03	0.49	57.76	-0.18	9.42	32.56	11.24	15.86	4	14.89	145.96
HC07	28.09	1.61	2.78	67.49	2.86	18.58	44.78	13.95	23.90	3	13.44	76.92
HC08	34.60	3.23	7.82	71.27	4.67	29.87	67.42	2.87	11.45	2	13.67	62.66
HC09	39.56	2.73	7.80	74.17	3.13	26.41	58.06	6.66	20.49	2	18.44	44.91
HC10	32.66	1.68	4.50	64.96	3.74	15.01	57.16	5.18	12.95	2	13.44	65.95
HC11	34.44	2.05	5.98	75.36	2.51	26.66	59.07	5.33	13.40	2	13.22	75.77
HC12	39.57	3.17	9.50	81.93	5.06	21.97	66.52	3.89	12.46	1	12.33	24.03
HC13	27.59	1.03	1.72	66.22	3.33	11.15	45.15	8.84	15.56	3	15.67	146.92
HC14 ^z	27.86	0.26	-0.83	70.09	-1.96	8.66	18.26	6.79	7.52	NE	18.67	215.14
HC15	33.01	1.60	2.59	74.84	3.54	15.35	54.28	5.19	12.36	2	14.00	87.83
HC16	40.81	8.20	13.84	81.05	11.13	36.18	62.45	6.88	21.89	5	13.00	24.14
HC17	37.93	3.19	7.69	74.97	3.2	34.34	62.21	3.98	12.44	1	14.00	41.04
HC18	32.35	1.88	4.26	72.02	3.9	13.55	45.30	10.93	20.35	2	14.56	86.59
HC19	42.18	3.86	10.75	83.55	3.89	34.33	67.04	3.47	12.15	5	12.00	30.83
HC20	32.79	1.96	4.81	73.47	5.2	17.61	55.97	5.48	13.49	2	11.89	90.48
HC21	57.06	4.27	16.23	88.1	8.13	19.37	74.80	0.15	6.72	1	8.56	22.63
HC22	42.21	4.38	10.87	88.86	11.1	21.44	67.78	3.29	11.86	1	13.67	31.04
HC23	59.02	3.29	13.74	94.02	4.62	18.75	72.61	0.65	7.45	1	11.44	26.90
HC24	31.82	2.87	1.20	69.65	5.15	6.21	45.98	7.20	12.88	3	13.67	118.56
HC25	42.82	7.22	11.12	82.23	17.45	26.05	66.86	4.41	17.30	5	12.78	42.31
HC26	31.21	2.66	1.42	71.25	6.45	6.61	38.52	9.46	15.55	2	13.67	118.84
HC27	33.03	3.14	2.47	71.75	7.38	5.86	44.51	9.38	16.94	2	16.00	101.59
HC28	32.56	2.97	1.56	65.47	7.2	4.18	39.07	15.11	23.16	2	14.78	103.35

HC29	39.33	3.48	6.93	78.48	8.83	17.31	60.75	5.58	15.89	2	12.11	38.74
HC30	31.67	2.10	2.02	58.13	7.67	6.22	41.81	13.40	21.68	3	14.44	114.41
HC31	36.63	2.29	4.70	77.8	10.42	14.5	58.27	8.19	22.13	2	12.22	75.32
HC32	32.50	2.66	4.99	67.63	8.57	13.34	45.98	15.23	26.94	2	16.00	57.11
HC33	79.05	-0.62	8.52	86.54	-0.02	17.69	78.43	-0.89	1.64	6		10.18
HC34	38.20	3.22	9.39	82.28	5.12	32.36	66.39	4.41	13.69	1	13.33	56.32
HC35	28.60	0.96	3.02	48.85	2.93	14.7	38.87	10.56	16.99	3	14.33	120.79
HC36	70.18	1.14	6.62	83.15	1.22	12.4	78.59	-1.48	2.04	6		10.23
MV0002	38.65	3.90	7.46	57.35	-0.96	11.37				1		
MV0004	29.17	0.95	1.91	50.12	3.99	1.87				3		
MV0006	31.11	1.91	1.70	71.79	5.59	5.7				3		
MV0007	34.62	3.29	2.41	70.48	6.81	10.52				2		
MV0008	30.92	2.13	3.43	44.37	8.71	-24.82				2		
MV0009	29.61	2.34	3.16	42.35	8.42	-27.93				2		
MV0010	28.10	0.74	0.11	39.18	8.75	-32.86				3		
MV0011	45.89	3.43	11.15	79.22	7.92	7.85				1		
MV0012	41.82	5.06	11.83	78.22	7.03	22.68				5		
MV0013	31.39	2.74	3.24	55.38	7.47	-2.16				2		
MV0014	40.02	3.54	7.48	74.22	8.46	13.29				1		
MV0015	30.71	1.90	2.11	63.35	7.63	-11.83				3		
MV0018	30.55	1.99	0.85	39.56	4.84	-4.83				3		
MV0019	30.19	1.16	0.82	43.71	3.6	-2.29				3		
MV0023	28.50	1.66	2.00	54.31	6.84	0.98				2		
MV0024	35.11	2.84	4.83	74.9	7.41	12.48				1		
MV0025	31.92	2.08	2.79	57.87	6.01	1.55				2		
MV0026	32.55	1.90	1.89	65.85	4.53	4.44				2		
MV0027	33.68	2.85	3.93	63.96	12.52	-7.6				2		
MV0028	31.40	2.09	0.27	61.98	6.3	-13.79				4		
MV0029	33.61	3.67	1.22	48.05	11.56	-24.82				3		
MV0030	39.15	5.19	6.74	72.06	15.6	0.34				1		
MV0031	32.91	2.97	1.97	69.61	12.61	19.71				2		
MV0032	31.87	2.84	0.87	69.75	9.61	7.25				3		

MV0033	46.87	5.21	11.19	84.88	6.82	19.07	1
MV0036	41.27	5.98	10.16	84	14.6	18.68	1
MV0037	34.57	2.21	2.91	23.6	2.04	26.75	3
MV0038	31.37	1.96	2.17	67.86	6.53	19.66	3
MV0040	27.94	0.90	0.64	58.42	2.23	13.14	4
MV0041	34.48	2.79	4.15	82.39	7.98	20.96	2
MV0042	35.30	4.17	7.03	69.3	10.85	27.99	5
MV0043	28.20	1.52	0.07	66.71	3.41	13.92	3
MV0044	43.47	4.59	10.66	89.03	5.79	25.43	1
MV0045	37.37	3.14	5.40	79.17	7.02	21.62	1
MV0046	44.70	4.37	9.55	97.59	6.29	25.43	1
MV0047	28.24	0.75	0.54	59.63	3.22	16.68	3
MV0048	39.75	10.67	15.67	80.66	19.92	44.44	5
MV0049	37.84	3.21	6.76	74.43	8.31	24.69	2
MV0050	32.42	1.98	3.17	75.19	7.24	14.43	2
MV0051	36.04	2.65	4.81	72.09	7.24	25.67	2
MV0052	32.57	2.54	2.54	58.72	7.74	17.59	3
MV0053	32.61	3.45	5.65	76.22	11.81	30.54	2
MV0054	57.42	4.04	14.14	88.33	1.85	23.38	1
MV0055	30.48	1.40	2.22	70.49	4.95	15.64	3
MV0056	30.44	2.08	2.14	75.88	6.58	18.65	3
MV0057	33.84	2.63	3.81	72.66	6.35	23	2
MV0058	35.92	4.00	6.99	71.88	5.97	22.8	2
MV0059	30.49	1.70	2.39	73.66	-2.79	29.15	2
MV0060	30.72	2.16	2.39	69.68	-2.73	25.54	3
MV0061	31.32	1.54	1.10	74.07	-1.94	24.16	3
MV0062	28.47	0.86	-0.04	65.26	3.26	16.76	3
MV0063	37.60	3.36	5.43	72.97	6.27	23.77	1
MV0064	30.77	2.30	0.25	70.04	5.33	17.57	3
MV0065	40.34	7.34	11.46	70.98	7.45	49.34	5
MV0066	51.81	4.70	11.54	81.68	2.8	29	1
MV0067	28.43	0.61	0.76	72.32	2.86	27.49	3

MV0068	29.69	1.71	1.50	69.42	-0.42	20.78	3
MV0069	28.83	1.76	1.80	76.82	8.67	45.8	3
MV0070	33.13	2.71	4.48	63.08	1.58	9.62	2
MV0071	52.34	4.20	13.57	92.1	-2.14	19.98	1
MV0074	37.32	3.89	0.80	71.92	8.09	-5.74	3
MV0075	38.11	3.70	0.42	63.46	8.72	-11.13	3
MV0076	52.90	5.80	11.60	82.93	15.33	7.43	1
MV0077	36.72	4.24	0.53	62.69	11.18	-11.99	3
MV0078	49.64	4.86	8.56	82.52	13.13	5.18	1
MV0079	44.45	4.95	5.27	80.72	13.05	3.74	1
MV0080	28.81	0.71	0.75	58.03	2.56	-0.84	3
MV0081	45.10	3.60	12.18	83.22	7.63	17.99	1
MV0082	37.28	3.18	7.32	72.14	11.38	16.75	2
MV0083	31.73	2.31	3.18	64.54	10.05	12.24	3
MV0084	31.13	1.67	1.30	51.35	3.47	0.4	3
MV0085	31.18	1.48	1.91	57.52	5.69	5.14	3
MV0086	32.74	2.68	4.83	67.71	8.29	10.14	2
MV0087	31.24	1.92	3.22	63.38	8.1	9.23	3
MV0088	42.49	3.78	10.85	79.18	9.28	19.57	1
MV0089	31.66	1.88	3.98	72.72	8.08	11.94	2
MV0090	28.62	1.92	2.19	57.45	6.65	6.47	3
MV0091	32.43	1.53	3.16	71.04	6.11	6.88	3
MV0092	35.51	2.77	5.89	69.32	7.98	11.79	2
MV0093	32.42	1.91	3.11	61.1	7.24	8.81	3
MV0094	32.12	2.56	3.84	64.46	8.03	9.77	2
MV0095	31.25	1.88	4.28	71.01	7.77	10.32	2
MV0096	27.64	1.15	0.75	50.5	3.97	1.27	4
MV0097	29.51	1.61	1.02	56.38	1.36	3.7	4
MV0098	42.91	4.76	11.50	83.85	8.51	24.12	1
MV0099	36.72	2.81	6.48	77.86	6.17	13.86	2
MV0100	30.34	1.50	2.43	71.4	5.86	8.72	3
MV0101	44.66	3.80	10.65	80.88	6.29	19.64	1

MV0102	37.47	2.79	5.61	75.41	5.13	9.36	2	
MV0103	32.20	2.56	4.97	68.43	7.27	14.72	2	
MV0104	48.55	4.28	12.82	82.63	4.97	20.31	1	
MV0105	27.06	-0.26	0.38	60.94	0.64	2.22	4	
MV0106	28.69	1.17	2.40	66.54	4.54	6.68	2	
MV0107	32.75	2.46	3.57	71.45	4.72	9.82	2	
MV0108	41.83	4.03	9.72	83.56	7.62	19.51	1	
MV0109	35.86	2.82	5.73	76.62	4.53	9.9	2	
MV0110	28.13	1.23	1.44	49.68	4.02	3.14	3	
MV0111	28.34	0.89	1.08	52.38	1.53	3.01	3	
MV0112	35.17	2.34	4.51	67.74	6.73	12.03	2	
MV0113	29.16	0.98	1.84	46.59	1.64	-0.12	4	
MV0114	29.24	1.73	2.63	48.9	6.22	5.77	3	
MV115	45.12	3.91	12.23	79.78	13.28	5.5	1	
MV116	33.10	3.51	0.40	61.61	3.71	8.65	3	
MV117	31.55	2.97	-0.27	54.36	1.18	5.15	3	
MV118	37.25	3.80	3.65	77.63	5.67	18.24	2	
MV119	32.02	3.36	0.05	56.65	1.54	19.5	3	
MV120	32.01	3.48	-0.09	54.11	0.85	16.94	3	
MV121	31.56	3.22	0.59	58.57	1.15	3.99	3	
MV122	49.67	4.76	11.09	76.48	6.24	23.4	1	
MV123	39.55	3.91	6.56	76.36	5.21	15.15	2	
MV124	32.83	1.96	4.63	70.78	5.83	11.69	2	67.85
MV125	42.67	3.79	8.70	78.27	7.8	16.91	1	31.71
MV126	39.11	3.23	4.55	70.65	10.58	15.39	1	51.07
MV127	38.17	3.50	5.75	78.25	7.35	14.42	2	79.39
MV128	36.24	7.21	13.34	79.53	17.75	30.03	5	24.54
MV129	35.33	2.41	5.69	68.5	10.54	15.87	2	
MV130	46.01	4.41	11.75	79.99	9.39	21.54	1	
MV131	29.09	1.26	-0.36	49.35	0.68	-1.47	3	162.80
MV132	28.92	1.35	0.00	53.15	1.06	1.05	3	112.27
MV133	30.81	2.31	1.70	49.38	4.31	4.51	3	

MV134	32.16	2.20	2.30	59.76	4.03	6.73	3	149.84
MV135	32.33	3.25	4.09	56.87	7.31	9.72	3	94.53
MV136	44.16	4.05	10.30	81.35	8.67	21.61	1	18.63
MV137	29.35	1.39	0.72	51.59	1.55	5.12	3	155.24
MV138	30.47	1.75	1.50	55.24	2.17	6.36	3	92.87
MV139	31.03	1.78	1.28	50.54	2.36	4.81	3	
MV140	34.83	3.39	5.07	54.91	7.08	11.32	3	94.89
MV141	37.33	3.66	6.89	74.27	7.32	19.9	2	35.52
MV142	28.81	1.48	0.97	51.72	2.32	4.09	3	96.01
MV143	42.53	7.12	13.48	78.1	15.58	27.54	5	
MV144	31.67	2.36	2.24	60.74	7.67	11.84	3	
MV145	30.07	2.25	-0.20	52.63	1.64	-9.83	4	
MV146	34.57	3.28	2.22	64.08	3.92	0.91	3	
MV147	35.95	3.84	2.99	59.75	5.88	-1.55	3	
MV148	37.97	4.52	7.38	67.12	10.82	9.78	5	
MV149	35.35	3.57	4.72	66.32	5.9	4.36	3	
MV150	38.23	4.37	7.82	73.72	8.45	11.85	2	
MV151	31.15	2.67	1.94	60.21	4.43	-1.36	3	
MV152	31.62	2.53	2.26	47.52	5.26	-11.87	3	
MV153	31.07	2.08	0.89	48.32	3.5	-12.05	3	
MV154	34.19	3.36	3.63	54.75	6.22	-3.9	3	
MV155	53.21	4.71	14.57	85.25	8.21	21.58	1	
MV156	31.27	1.65	2.29	54.59	4.71	-5.82	3	
MV157	32.90	2.54	4.35	54.47	6.06	-4.15	2	
MV158	30.31	1.35	1.44	45.91	4.76	-15.28	3	117.74
MV159	35.11	3.28	4.78	49.95	7.5	-6.59	2	75.49
MV160	31.19	3.40	3.14	56.52	6.7	-4.84	3	62.78
MV161	30.07	2.10	2.25	48.3	4.65	-10.05	3	
MV162	28.67	0.59	0.57	49.85	2.22	-10.1	3	
MV163	32.47	2.57	4.95	55.56	6.23	-3.21	2	
MV164	31.37	0.89	1.45	53.56	2.85	-6.01	3	
MV165	32.02	1.76	3.17	57	4.11	-2.91	3	

MV166	30.66	1.63	1.37	46.66	6.95	-15.53		3	
MV167	34.63	4.72	6.97	64.03	9.33	3.52		2	
MV168	37.78	5.44	10.59	61.73	12.96	6.99		5	38.01
MV169	34.67	2.31	4.23	63.48	6.64	3.1		2	
MV170	30.55	1.68	0.43	48.39	2.87	-12.05		3	
MV171	38.09	3.01	7.13	73.07	6.22	9.63		2	
MV172	39.04	8.51	13.90	66.72	19.12	12.29		5	17.63
MV173	35.12	10.07	9.53	53.35	20.89	-1.22		5	30.60
MV174	33.33	2.41	4.03	53.24	5.85	-6.19		3	
MV175	40.52	4.15	9.19	73.64	7.37	12.59		1	47.89
MV176	31.04	2.16	1.35	53.32	3.92	-8.75		3	
CR04 ²	27.97	-0.63	-0.72	46.91	-0.39	-9.08			
CR05 ²	25.82	-0.87	-0.74	53.49	-2.53	8.39			
CR07 ²	22.84	-0.60	-0.86	41.98	0.46	-14.02			

¹Observer Reported Colours: 1 = Blonde 2 = Light Brown 3 = Dark Brown 4 = Black 5 = Red 6 = White

²Non-European Individuals

APPENDIX 2 – Computer Code

aRGB2LAB Conversion Module for V++

{Images need to be .jpg in the following folders:

- folder containing subfolders:
 - images
 - white balance images

The images for testing are saved to the 'images' folder.

The white balance images are saved in the 'white balance images' folder.

The name of the 'white balance image' must be the same as the 'image' with a 'w' appended as a suffix.}

```
button btn_text,'Lab stats from folder w/ white correction';
var
img,ref_present, white_ref;
temp_small,temp_big;
img_red,img_green,img_blue;
ref_red,ref_green,ref_blue;
l,la,lb;
rl,ra,rb;
t_small, t_big;
name, imgname,whiterefname, path;
ed,ed2;

const//***** white
level values
white_l=93.4;
white_a=1.16;
white_b=-2.06;

function BrowseFolders( Path:pointer ) : integer ; external
'VppCommonDlgs.dll' ;
function GetFolder ;
var
FolderName ;
begin
FolderName := StringOfChar( ' ',255 ) ;
if BrowseFolders( FolderName ) <> 0 then
GetFolder := Trim( FolderName )
else
GetFolder := ' ' ;
end; GetFolder }

function powerof(s,t);//***** get 's' to
the power 't'
var
value,result;
begin
value:=Ln(s);
result:=Exp(value*t);
powerof:=result;
end;

function get_channel(i);//***** get RGB
channel
var
t_small,t_big,c;
begin
c:=single(i)/255;
t_small:=c*(c<=0.04045)/12.92;
t_big:=powerof(((c+0.055)/1.055),2.4)*(c>0.04045);
get_channel:=t_small+t_big;
```

```

end;

function get_l (i,r,g,b); //***** get
L channel
var
yn,y,yy,temp_big,temp_small;
begin
yn:=0.212671+0.715160+0.072169;
y:=0.212671*r+0.715160*g+0.072169*b;
yy:=y/yn;
temp_big:=( 116*powerof(yy,0.3333)-16)*(yy>0.008856);
temp_small:=yy*(yy<=0.008856)*903;
get_l:=temp_big+temp_small;
end;

function get_a (i,r,g,b); //***** get
a channel
var
xn,yn,x,y,xx,yy,temp_big,temp_small;
begin
xn:=0.412453+0.357580+0.180423;
yn:=0.212671+0.715160+0.072169;
x:=0.412453*r+0.357580*g+0.18042*b;
y:=0.212671*r+0.715160*g+0.072169*b;
xx:=x/xn;
yy:=y/yn;
temp_big:=powerof(xx,0.3333)*(xx>0.008856);
temp_small:=(7.787*xx+0.1379310345)*(xx<=0.008858);
xx:=temp_big+temp_small;
temp_big:=powerof(yy,0.3333)*(yy>0.008856);
temp_small:=(7.787*yy + 0.1379310345)*(yy<=0.008856);
yy:=temp_big+temp_small;
get_a := 500*(xx-yy);
end;

function get_b (i,r,g,b); //***** get
b channel
var
yn,zn,y,z,yy,zz,temp_big,temp_small;
begin
yn:=0.212671+0.715160+0.072169;
zn:=0.019334+0.119193+0.950227;
y:=0.212671*r+0.715160*g+0.072169*b;
z:=0.019334*r+0.119193*g+0.950227*b;
yy:=y/yn;
zz:=z/zn;
temp_big:=powerof(yy,0.3333)*(yy>0.008856);
temp_small:=(7.787*yy + 0.1379310345)*(yy<=0.008856);
yy:=temp_big+temp_small;
temp_big := powerof(zz,0.3333)*(zz>0.008856);
temp_small := (7.787*zz + 0.1379310345)*(zz<=0.008856);
zz := temp_big + temp_small;
get_b := 200*(yy-zz);
end;

begin //***** program
start
path:=getfolder;
Name := FindFirstFile( path+'\images\*.jpg',fa_Archive );
ed:=createeditor('output');
ed2:=createeditor('corrected mean values');
writeln(ed2,chr(9)+'L'+chr(9)+'a'+chr(9)+'b');
while StrLen( Name ) > 0 do

```

```

begin
imgname:=name;
open (imgname, img);
converttype (img, typ_rgbfloat);
img:=img+0.01;
//writeln (imgname);
whiterefname := path+' \white balance
images\'+'w'+extractfilename (imgname) ;
open (whiterefname, white_ref);
//writeln (whiterefname);

img_red:=get_channel (red (img));
img_green:=get_channel (green (img));
img_blue:=get_channel (blue (img));

l:=get_l (img, img_red, img_green, img_blue); //*****
***** get image Lab channels
la:=get_a (img, img_red, img_green, img_blue);
lb:=get_b (img, img_red, img_green, img_blue);

//***** write
image details
writeln (ed, '_____');
writeln (ed, 'Image for conversion is:');
writeln (ed, extractfilename (imgname)+chr (9)+'L'+chr (9)+'a'+chr (9)+'b');
writeln (ed, 'Mean'+chr (9)+str (meanof (l))+chr (9)+str (meanof (la))+chr (9)+
str (meanof (lb)));
writeln (ed, 'Deviation'+chr (9)+str (stdof (l))+chr (9)+str (stdof (la))+chr (
9)+str (stdof (lb)));
writeln (ed);

//***** get white
reference Lab channels

ref_red:=get_channel (red (white_ref));
ref_green:=get_channel (green (white_ref));
ref_blue:=get_channel (blue (white_ref));

rl:=get_l (white_ref, ref_red, ref_green, ref_blue); //show (rl);
ra:=get_a (white_ref, ref_red, ref_green, ref_blue); //show (ra);
rb:=get_b (white_ref, ref_red, ref_green, ref_blue); //show (rb);
//***** write
whitebalance image details
writeln (ed, 'White reference image is:');
writeln (ed, extractfilename (whiterefname)+chr (9)+'L'+chr (9)+'a'+chr (9)+
'b');
writeln (ed, 'Mean'+chr (9)+str (meanof (rl))+chr (9)+str (meanof (ra))+chr (9)
+str (meanof (rb)));
writeln (ed, 'Deviation'+chr (9)+str (stdof (rl))+chr (9)+str (stdof (ra))+chr
(9)+str (stdof (rb)));
writeln (ed);
//***** write
corrected image details
writeln (ed, 'Corrected image values:'); //***** write
corrected image details
writeln (ed, 'Mean'+chr (9) +str ((white_l-meanof (rl))+meanof (l))+chr (9)
+str ((white_a-meanof (ra))+meanof (la))+chr (9) +str ((white_b-
meanof (rb))+meanof (lb)));
writeln (ed2, extractfilename (imgname)+chr (9) +str ((white_l-
meanof (rl))+meanof (l))+chr (9) +str ((white_a-
meanof (ra))+meanof (la))+chr (9) +str ((white_b-
meanof (rb))+meanof (lb)));

```

```
Name := FindNextFile ;  
end;  
  
copytoclipboard(ed2);  
  
writeinfo('complete');  
  
free (img);  
free (l);  
free (la);  
free (lb);  
free (rl);  
free (ra);  
free (rb);  
end
```

**APPENDIX 3 – Project Information Sheet, Consent Forms
and Questionnaire**



Blond Hair Colour: Classification and Genetic Associations

Background Information

This study is conducted at Victoria University by PhD candidate, Michelle Vaughn and her supervisor, Dr. Swati Baidur-Hudson (03 9216 8123) with the support of Dr. Roland van Oorschot (Associate Investigator and Co-supervisor) from the Victoria Police Forensic Services.

Researchers all over the world are developing tests for physical features such as eye colour and red hair. This study is concerned with characterising hair colour and looking at genetic variations that may contribute to blonde or fair hair. To do this we need to examine the hair at many levels and look at a persons DNA to find the correlations.

If the research is successful, the ability to describe an individual from their DNA has many potential uses:

It may be possible in the future to determine some of a person's physical features from DNA inadvertently left at a crime scene. This would allow the police to use this information as an investigative tool for limiting a suspect pool, or to provide a starting point for a search, much as geographical or psychological profiling are used now in many jurisdictions. Other circumstances where it would be useful to be able to predict the physical features of a person from their DNA include natural and criminal mass disasters, where the appearance of victims may have been destroyed beyond recognition. If DNA could be recovered from these victims and some of their physical features predicted it may make it easier for forensic investigators to narrow the search for their identity.

The procedures and samples required for this study are listed on the back of this information sheet. You may choose to participate in some or all of them, at your discretion.

While this research may one day benefit forensic scientists and investigators, your samples are going to be used purely for research, your identity kept anonymous, and information about you or your DNA sample will be kept totally confidential.

Steps taken to ensure privacy, anonymity and confidentiality of volunteers

Ethics approval has been obtained from the Victoria University Human Research Ethics Committee, approval number HRETH 06/156. Ethics approval is necessary to protect the rights of the people taking part in this study. The conditions of the approval include:

- Samples will only be taken from volunteers with informed consent.
- Each volunteer has the opportunity to withdraw from the project at anytime.
- All details recorded will be kept in strictest confidence, under lock and key and only the researchers will have access to any information given while respecting the requirements of the law.
- Samples and questionnaire answers will only be identified by code number not by name.
- To ensure the privacy of volunteers, data will only be presented as linked to physical characteristics such as hair colour, and not related to individuals
- Neither your name nor contact details will be linked to samples or data generated from analysis. The data will only be linked to physical characteristics.
- All names and contact details will be destroyed at the end of the study
- At the completion of the entire study, all details as well as any remaining swabs and extracted DNA samples will be destroyed.

To ensure your safety, privacy and confidentiality at all times, we have taken a number of measures:

Physical: The sampling method used will be the collection of inner cheek cells with sterile swabs, which will only be taken by you. This will make the procedure quick, easy and least invasive for you. If hair is to be plucked and clipped, all efforts will be made to reduce discomfort.

Psychological/social/legal: Any data collected from the answers to the questionnaire or from the analysis of samples will be used for research purposes only and will not be made available to any other party. Strict confidentiality and security measures will be maintained – see details above.

Procedures:

- After reading the above information, you will be asked if you want to participate in the study.
- If so, you will be asked to fill out and sign the consent form
- You will be assigned a code and from there on your samples will only be referred to with the code to preserve your anonymity.
- You will then be asked to provide a sample of inner cheek cells (buccal cells) using sterile swabs.
- If you have given permission, you will be requested to fill out a questionnaire about some of your physical characteristics, which will only be identified by your assigned code, no personal details
- If have given permission, a digital photo will be taken of your hair for colour determination. Measurements of your hair colour will also be taken with a reflective spectrophotometer. This instrument uses normal light to measure colour. Again, all these will be recorded against the code only.
- If you have given permission, several hairs will be plucked from your scalp and approximately 30 mg (a small amount) of hair will be clipped.
- The collected samples will be stored and may be used to study links between genetic variations and the physical characteristics of hair colour.

Sample code: _____

Adult Consent Form

I, _____ of
(address) _____

certify that I am at least 18 years old and that I am voluntarily giving my consent to participate in the experiment entitled:

Blonde Hair Colour: Classification and Genetic Associations

being conducted at Victoria University, by PhD candidate, Michelle Vaughn and her supervisor, Dr. Swati Baidur-Hudson, with the support of Dr. Roland van Oorschot (Associate Investigator and Co-supervisor) at the Victoria Police Forensic Services.

I certify that the objectives of the experiment, together with any risks associated with the procedures to be carried out in the experiment, have been fully explained to me by

(Name of researcher) _____ on (date) _____

(Signature of researcher): _____

and that I have had the opportunity to have any questions answered. I also understand that I can withdraw from this experiment at any time and that this withdrawal will not jeopardise me in any way. I have been informed that the information I provide will be kept confidential.

I freely consent to participation in the following procedures:

Having a digital photo and other colour measurements taken of myself to help determine the exact colour of my hair..... Yes No

Filling out of a questionnaire regarding some of my physical characteristics..... Yes No

Self-taking of buccal cell samples (for DNA) using sterile swabs..... Yes No

Plucking of several scalp hairs..... Yes No

Trimming of approximately 30 mg of scalp hair Yes No

To be contacted if need be in future, should a sample need to be repeated..... Yes No

If yes, your telephone no.: _____ and/or email address: _____

Signature: _____ **Date:** _____



Sample code: _____

Revocation of Consent Form

I, _____ of
(address) _____

hereby wish to WITHDRAW my consent to participate in the research project

Blonde Hair Colour: Classification and Genetic Associations

being conducted at Victoria University, by PhD candidate, Michelle Vaughn and her supervisor, Dr. Swati Baidur-Hudson, with the support of Dr. Roland van Oorschot (Associate Investigator and Co-supervisor) at the Victoria Police Forensic Services.

I understand that such withdrawal WILL NOT jeopardise any treatment or my relationship with Victoria University or the Victoria Police Forensic Science Services (if applicable)

Any data already collected may / may not be included in the research project (circle one).

Signature _____ **Date:** _____

Sample Code: _____

Researcher's initials _____



**VICTORIA
UNIVERSITY**

**A NEW
SCHOOL OF
THOUGHT**

Date completed: ____ / ____ / ____
(Day / Month / Year)

General Information

1. Reason for interest in project:	2. Date of Birth: ____ / ____ / ____ (Day / Month / Year)	3. Gender: Male <input type="checkbox"/> Female <input type="checkbox"/>
4. Do you currently have your natural adult hair colour? <input type="checkbox"/> Yes <input type="checkbox"/> No	5. Are you taking any medication or nutritional supplements that have altered your adult hair colour? <input type="checkbox"/> Yes <input type="checkbox"/> No	

Pigmentation Characteristics

		Or describe in your own words:
Hair Colour	Black <input type="checkbox"/> Dark Brown <input type="checkbox"/> Light Brown <input type="checkbox"/> Blonde <input type="checkbox"/> Red <input type="checkbox"/>	
Eye Colour	Black <input type="checkbox"/> Dark Brown <input type="checkbox"/> Light Brown <input type="checkbox"/> Green <input type="checkbox"/> Blue <input type="checkbox"/> Grey <input type="checkbox"/> Hazel <input type="checkbox"/>	

Background (please tick appropriate box for each)

Ethnicity*:	Self	Biological Mother	Biological Father	Maternal Grandmother	Maternal Grandfather	Paternal Grandmother	Paternal Grandfather
North European							
South European							
Middle Eastern							
South Asian							
Oriental Asian							
Pacific Islander							
Australian aboriginal							
African Negroid							
Native American							
Other :							

* **North European:** France, Switzerland, Belgium, Netherlands, Denmark, Germany, Austria, Czech Rep., Slovakia, Hungary, Romania, Poland, Ukraine, Moldova, Belarus, Lithuania, Latvia, Estonia, West Russia, Iceland, Greenland (Denmark), United Kingdom, Sweden, Norway, Finland **South European:** Portugal, Spain, Monaco, Italy, Malta, Slovenia, Croatia, Bosnia and Herzegovina, Serbia and Montenegro, Greece, F.Y.R.O.M (Macedonia), Bulgaria, Turkey

Researcher's Observations:

Pigmentation Characteristics

		Or describe in your own words:
Hair Colour	Black <input type="checkbox"/> Dark Brown <input type="checkbox"/> Light Brown <input type="checkbox"/> Blonde <input type="checkbox"/> Red <input type="checkbox"/>	
Eye Colour	Black <input type="checkbox"/> Dark Brown <input type="checkbox"/> Light Brown <input type="checkbox"/> Green <input type="checkbox"/> Blue <input type="checkbox"/> Grey <input type="checkbox"/> Hazel <input type="checkbox"/>	

Notes:

APPENDIX 4 – Published Papers

Hair Color Measurement and Variation

Michelle Vaughn,^{1*} Roland van Oorschot,² and Swati Baidur-Hudson¹

¹*School of Molecular Sciences, Victoria University, Melbourne, VIC Australia*

²*Victorian Police Forensic Services Department, Macleod, VIC Australia*

KEY WORDS hair pigmentation; reflective spectrophotometry; cluster analysis; forensic science

ABSTRACT Pigmentation of hair in humans has been investigated by medical scientists, anthropologists and, more recently, by forensic scientists. In every investigation, hair color must first be defined by the researchers. Subjective color assessment inhibits the reproducibility of experiments and the direct comparison of results. The aim of this study was to objectively measure human hair color and examine the variation found in a population with European ancestry, using the CIE L*a*b* color space. Observer-perceived hair colors were compared with self-reported hair colors and the color as measured by reflective spectrophotometry of 132 subjects of European ancestry. The presented data show that self-reported hair

colors and observer-reported colors are similar; however, these categories are not necessarily the best way to categorize hair color for quantitative research. Using a two-step cluster analysis, hair color can be divided into categories or clusters based on spectrophotometric measurements in the CIE L*a*b* color space and these clusters can be well discriminated from each other. This separation is primarily based on the b* (yellow) color component and the clusters show agreement to observer-reported colors. This study illustrates the possibilities for and necessity of objectively defining the hair color phenotype for various downstream applications. *Am J Phys Anthropol* 137:91–96, 2008. © 2008 Wiley-Liss, Inc.

Human pigmentation has been observed and studied for over 4000 years (Westerhof, 2006) and is an integral part of our physical health and cultural identity. Normal pigmentation variation in humans has been investigated by medical scientists (Duffy et al., 2004), anthropologists (Parra et al., 2004; Parra, 2007) and, more recently, by forensic science researchers for the prediction of physical characteristics from DNA, to be used as an investigative tool. The prediction of physical traits for use in forensic investigations is a developing field and hair color is one of many traits that may assist investigators in limiting a suspect pool. The UK Forensic Science Service has already implemented a test with 84% accuracy for predicting red hair (Grimes et al., 2001).

Hair color has received somewhat less attention in the study of pigmentation than skin or eye color, with most work on hair color being done as a consequence of, or in conjunction with, other pigmentary traits (such as skin color or melanoma risk) (Sturm et al., 2003; Duffy et al., 2004). These investigations have led to a great deal of information on red hair (Box et al., 1997; Grimes et al., 2001; Ha et al., 2003; Sturm et al., 2003; Duffy et al., 2004; Naysmith et al., 2004), which is associated with skin cancer risk; however data on non-red hair colors is scarce. Researchers contributing to the scientific investigation of hair color represent the different fields of medical science and genetics (Ancans et al., 2003; Ha et al., 2003; Duffy et al., 2004), the cosmetics industry (Takahashi and Nakamura, 2004, 2005), and forensic science (Grimes et al., 2001), as well as the disciplines of chemistry and toxicology (Borges et al., 2001; Nogueira and Joekes, 2004). There has also been interest in the hair color of our hominid relatives; a recent study has determined that Neanderthals had variants in a gene that is associated with differing pigmentation levels in *Homo sapiens* (Lalueza-Fox et al., 2007).

The extinct Neanderthals were a hominid species that lived in Europe and western Asia (Krause et al., 2007).

Modern *Homo sapiens* from the same geographical area show almost all of the hair color variation found in human populations. From studies of genetic sequence diversity (Harding et al., 2000), it has been concluded that there was positive selection pressure for maintaining dark skin and hair pigmentation in the southern latitudes of Africa and Asia, where it protected an individual from ultraviolet light from the sun (Rees, 2003; Parra, 2007). When human populations migrated to more northern latitudes, there was less pressure to maintain dark pigmentation and instead, light skin pigmentation became a selective advantage (e.g., through the easier synthesis of Vitamin D) (Rees, 2003; Parra, 2007). The possibility of sexual selection being a factor in the evolution of pigmentation variation is supported by the presence of sexual dimorphism in skin pigmentation and by evidence suggesting that pigmentation influences mate selection (Aoki, 2002; Madrigal and Kelly, 2007a; Parra, 2007); however, difficulties arise in trying to isolate and quantify the effects of natural and sexual selection (Frost, 2007; Madrigal and Kelly, 2007a,b).

Evidence suggests that hair and skin color are somewhat correlated, within the European population (Shriver and Parra, 2000), and hair color is likely the result of similar selective pressures; however, the exact genetic mechanisms of hair color and its high variability, as well as other factors in the biology of hair color have not yet been fully elucidated.

*Correspondence to: Michelle Vaughn, School of Molecular Sciences, Victoria University, Hoppers Lane, Werribee VIC 3030, Australia. E-mail: michelle.vaughn@live.vu.edu.au

Received 11 December 2007; accepted 11 March 2008

DOI 10.1002/ajpa.20849

Published online 29 April 2008 in Wiley InterScience (www.interscience.wiley.com).

In every investigation, "hair color" must first be defined by the researchers. What one investigation calls "blonde" however, might be what another calls "light brown" and this inhibits the reproducibility of experiments and the direct comparison of results.

Reflective spectrophotometry has been used to objectively measure skin pigmentation (Shriver and Parra, 2000; Parra et al., 2004) and the technique has also been used on hair in different ways and for various purposes including cosmetic science (Takahashi and Nakamura, 2004, 2005), for the determination of hair composition and structure (Nogueira and Joeke, 2004), for comparison of measurement methods (Shriver and Parra, 2000), and in recent genetic studies (Naysmith et al., 2004). It has been found to be an objective and effective way to measure hair color.

To objectively measure hair color presupposes a method of color measurement. There are, however, numerous models or "color spaces" used to describe color. The model developed by the Commission Internationale de l'Éclairage, CIELAB or CIE $L^*a^*b^*$, measures color on three axes that are nearly linear with human perception (Ford and Roberts, 1998). This model provides a grid point for each specific color (TASI, 2004) and in addition, has been used for other studies of human pigmentation (Shriver and Parra, 2000; Takahashi and Nakamura, 2004, 2005; Parra, 2007).

In the CIE $L^*a^*b^*$ color space, the lightness, or intensity, of a color is measured on the " L^* " axis on a scale from 0 (black) to 100 (white). The color is then measured on the " a^* " axis which gives a value from -100 (green) to $+100$ (red). The " b^* " axis measures color from -100 (blue) to $+100$ (yellow). One unit on the L^* , a^* , or b^* axes is considered to be the smallest difference the human eye can detect (TASI, 2004). This color grid point allows for the mathematical comparison of color. It should be noted that, theoretically, the a and b axes have no maximum or minimum values but this research has used the cut off points of ± 100 because these represent the practical limit of the instrument used in the color measurement (Napier S, personal communication, Biolab Group, Australia, 2007).

In 2000, Shriver and Parra used the CIE $L^*a^*b^*$ system with reflective spectrophotometry to measure the hair color of 41 European-American individuals and 18 individuals of non-European ancestry, comparing the L^* color component to the Melanin Index. They found a significant correlation between these systems of measurement and reported low variability in the hair color of individuals of non-European ancestry. Naysmith et al. in 2004 used the CIE $L^*a^*b^*$ system to measure the hair color of 50 individuals for genetic and chemical studies of red hair. They found relationships between variations in the gene MC1R and measured hair color (most strongly with b^*), as well as with chemical studies. It should be noted that people with red hair were deliberately over-represented in this sample and volunteers ranged in age from 6 to 72 years old (median of 35). Both studies emphasized the necessity and demonstrated the feasibility of measuring pigmentation objectively and accurately.

The aim of the research presented here was to objectively measure human hair color and examine the variation found in a population with European ancestry, over all three color axes. Young adult volunteers were recruited due to the fact that hair color is known to change with age and especially during puberty and late

adulthood (Slominski et al., 2004). We were particularly interested in the variation found among the non-red colors (i.e., blonde, brown, and black). Observer-perceived hair colors were compared with self-reported hair colors and the color as measured by reflective spectrophotometry. The variation of the measured hair colors was then analyzed to determine if and how hair colors form describable groups and how these match with our intuitive descriptions of hair color.

Objectively describing categories of hair color will benefit the fields of anthropology and medical science and may introduce some standardization into the design of genetic studies. It could also benefit forensic investigation on various levels, for example, specific criteria may help in better communication between witnesses, investigators, and forensic hair-examiners and be an important part in the future of forensic phenotype-profiling and prediction of physical features from DNA samples.

MATERIALS AND METHODS

Subjects

Subject recruitment and sampling procedures were conducted with the approval of the Victoria University Human Research Ethics Committee. In total, 140 subjects were included in this analysis. All volunteers recruited for the study had their natural hair color at the time of sampling. Most volunteers were between the ages of 18 and 35 and were of European ancestry (132 subjects). The exceptions to these criteria include six individuals of non-European ancestry (one African and five South Asians) and two older European individuals with white hair. Subjects fitting these criteria were chosen to meet the study aims of examining natural variation in adult European hair color, with non-European and mature-white haired individuals included for comparison purposes.

Sampling

Subjects were first given all project information and the following procedures carried out following their written consent (in compliance with Victoria University Human Research and Ethics standards; approval number HRETH06/156). Subjects filled out a questionnaire that, in addition to confirming their age, ancestry, and natural hair color status, asked them to report how they saw their own hair and eye color. Hair color was reported to be one of; Black, Dark Brown, Light Brown, Blonde, or Red. Color assessment by an observer (the researcher taking the samples) was also recorded.

Hair color was measured by reflective spectrophotometry (RS) using a Minolta CR-300 Chroma Meter (Konica Minolta, North Ryde NSW, Australia). The instrument was calibrated using a white tile and a light source input setting of "D65" which represents daylight without spectral highlights. The instrument was set to measure in the CIE $L^*a^*b^*$ format. Hair color was measured on the left, right, and back of the head and was measured five times at each spot, the mean of the 15 measurements being the focus of subsequent statistical considerations. The standard deviation of the 15 measurements for each individual ranged from 0.07 to 4.67 for L^* , 0.01 to 2.51 for a^* , and 0.03 to 3.3 for b^* with means of 1.22, 0.33, and 0.68, respectively.

TABLE 1. The self-reported hair colors (SRC) compared to the observer-reported hair colors (ORC)

Color	Self reported	Observer reported
Black	8	12
Dark Brown	52	44
Light Brown	37	44
Blonde	33	31
Red	8	7
White	8	12

Statistics

SPSS 15.0 for Windows (©SPSS Inc. 2006, Chicago IL) was used to analyze the color data. Two separate methods were used—a cluster analysis and discriminant analysis. Measured hair color (by RS) was broken into several groups using a two-step cluster analysis; performed using the Euclidian distance criterion or the log-likelihood probability approach and either not specifying the number of clusters to be formed or asking for a specific number of clusters, ranging from two to seven. Cluster analysis seeks to identify natural subgroups within a population by minimizing within-group variation while maximizing between-group variation. To characterize and evaluate the clusters a separate, discriminant analysis was then performed. Discriminant analysis commences from the perspective of there being a known number of population subgroups. From the population, there are a number of individuals whose subgroup classification is known. The data from these individuals is analyzed in an attempt to build a profile of subgroup membership and subsequently use this profile to classify new individuals (whose subgroup membership is otherwise unknown).

After the analysis, SPSS determines which variables are most important in discriminating the groups, and how good this model will be at predicting group membership of future individuals. It determines the predictive value of the model by removing each individual separately, reanalyzing the data, and then predicting the membership of that case. The percent correctly classified is reported.

RESULTS

Reported colors

Self-reported hair colors (SRC) were compared to the colors determined by the observer (observer-reported color or ORC) (Table 1). 85.7% of individuals had the same hair color reported by themselves and the observer. Where there was a difference, discrepancies were lighter or darker by one shade. The observer was more likely (55% of disagreeing observations) to see a darker shade. The observer classified four SRC Dark Browns as Black, four SRC Light Browns as Dark Browns, and three SRC Blondes as Light Brown. The observer also classified eight SRC Dark Browns as Light Browns. In one case, a self-reported Red color was reported as Blonde by the observer.

Population variation

To examine the hair color variation amongst the European population, the L^* , a^* , and b^* values of the 134 European individuals, as measured by reflective spectro-

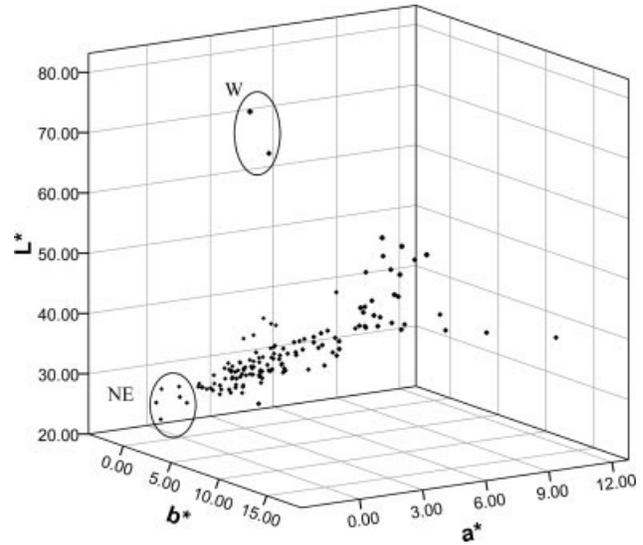


Fig. 1. Population Sample; the non-European samples (NE) and the mature-white haired samples (W) are highlighted. The scales for the CIE $L^*a^*b^*$ axes are $L^* = 0$ (Black) to 100 (White), $a^* = -100$ (green) to +100 (red), and $b^* = -100$ (blue) to +100 (yellow). Only the relevant portions of these scales are shown.

photometry, were recorded and the data subjected to statistical analysis. The variability observed in the entire sample is illustrated in Figure 1. This distinctiveness of the non-European individuals (NE in Fig. 1) and the individuals with white hair (W in Fig. 1) can be seen in the highlighted areas. Individuals not of European ancestry have hair that is much darker with a^* and b^* values very close to zero.

Grouping the population

The aim of the statistical analysis was to investigate the major components of the color variation and to see whether hair color falls into groups that would be meaningful for further research. What qualifies as a meaningful group is subjective and may depend on the intended use of the data; however in this case, meaningful groups would be those that separate colors into several distinct groups. For practical purposes, it would also be useful to have groupings that correspond well to colors people already use to report hair color (e.g. Blonde).

The European sample was broken into groups using the two-step cluster analysis. The analysis was first conducted without specifying the number of clusters, using either the Euclidian distance approach or the log-likelihood probability approach. Next, the numbers of clusters were specified, from two to seven for both approaches and a discriminant analysis was then performed. Using the clusters determined as a definitive population subgroup, the percentage of cases correctly classified was used to evaluate the clusters.

The percentage of cases correctly classified in the discriminant analysis for the various methods of cluster determination is shown in Figure 2. When the observer-reported color (ORC) groupings are used, only 73.1% of cases are correctly classified. All other clusters have greater than 95% of cases correctly classified.

Determined automatically, without specifying the number of clusters, the Euclidian distance approach

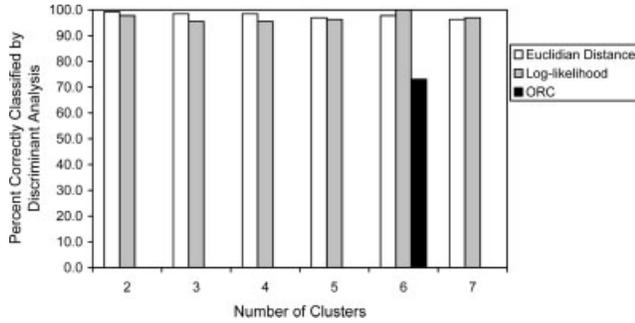


Fig. 2. Discriminant scores for each number of clusters and the method by which they were determined.

TABLE 2. Cluster distribution and characteristics for two clusters determined automatically by the log-likelihood approach (Analysis 1)

	N	% Of Total	L*		a*		b*	
			Mean	SD	Mean	SD	Mean	SD
Cluster 1	35	26.1%	47.05	8.95	4.55	1.88	10.78	2.45
Cluster 2	99	73.9%	32.15	3.08	2.16	0.90	2.99	2.05
Total	134	100%						

yielded no useful clusters, while the log-likelihood method grouped the population into two clusters, the composition of which are shown in Table 2 as an example of how clusters are described. When using the Euclidian distance to determine clusters, having two or three clusters results in very poor sample composition, with more than 97% and 93.3% of cases, respectively, in one cluster. From the remaining groups, two and six clusters as determined by the log-likelihood approach (which will be referred to as Analysis 1 and 2, respectively), and four clusters as determined by the Euclidian distance approach (which will be referred to as Analysis 3) showed the highest discriminant scores and the best group composition (data not shown).

During analysis, canonical discriminant functions are determined. These unstandardized functions are used to make classifications in the discriminant analysis and the standardized function coefficients are used to compare the relative importance of each function and each variable in making classifications. The coefficients of the standardized functions shown in Table 3 show that the most informative function in all three analyses is most highly correlated with the b* component of color (yellowness) and where there is more than one function, the second most important function is most highly correlated with L* (lightness). In Analyses 2 and 3, the third discriminant function contributed 1% and 0.9% of information, respectively (data not shown).

An example on the use of the unstandardized functions is shown in Figure 3, which illustrates the use of the function in discriminating between the two clusters determined in Analysis 1 ($F = -5.615 + 0.102L^* + 0.212a^* + 0.270b^*$). Using this function, 97.8% of new cases are correctly classified and this clear distinction between the two groups is visible in Figure 3.

Comparing clusters to the observer-reported colors

It appears that the clusters determined by various methods do show similarity to hair colors as perceived

TABLE 3. Coefficients of the standardized discriminant functions for the three discriminant analyses

Analysis function	1		2		3	
	1	1	2	1	2	
L*	0.535	0.214	0.988	-0.506	1.230	
a*	0.261	0.586	-0.613	-0.612	-0.405	
b*	0.585	0.754	-0.294	0.945	-0.228	
	100%	76.3%	22.6%	57.0%	42.1%	

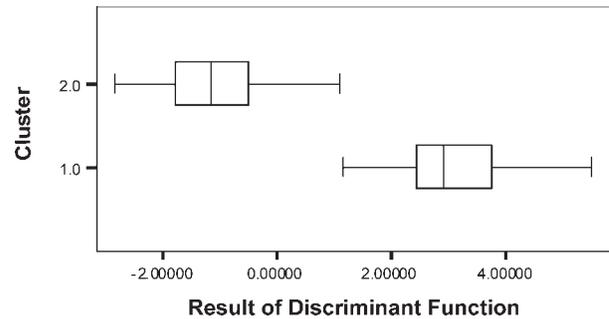


Fig. 3. The results of applying the discriminant function on the L*a*b* data, divided by cluster for the two clusters determined by the log-likelihood method (Analysis 1).

by an observer. When divided into two clusters by the log-likelihood approach (Analysis 1), the observer-reported hair colors Black, Dark Brown, Light Brown, and the four darkest Blondes are in Cluster 1 and the remaining Blondes and all Reds and Whites are in Cluster 2. This shows that the separation in these clusters determined by their L*, a*, and b* values correspond well to divisions in observer-reported colors and could be referred to as “Dark” and “Fair.” The names are arbitrary; however, they correspond to how people tend to refer to hair color.

The six clusters determined by log-likelihood (Analysis 2) separate the ORC White very clearly and the four Reds with the highest a* value (the most red). The other colors are somewhat less distinct; however, they are roughly divided into categories that could be called “Fair,” “Light,” “Medium,” and “Dark.” The example shown in Figure 4 illustrates the separation of the six clusters by applying the two most important discriminant functions (a) and the results of using the same functions on the observer-reported groupings (b), where blurring between the groups can be seen.

When four clusters are determined by the Euclidian Distance approach (Analysis 3), the ORC White is clearly separated, as are the two Reds with the highest a* and b* values. The rest of the population is divided into two clusters, the first with the darker colors and the other with the fair colors. It should be noted that the separation of fair from dark hair in these groups is similar to that of the two cluster groupings, with four extreme individuals (very light and yellow and the very red) being separated.

DISCUSSION

The presented data show that when given a limited choice of colors, self-reported hair colors and those reported by an observer are fairly consistent; however, these categories are not necessarily the best way to cate-

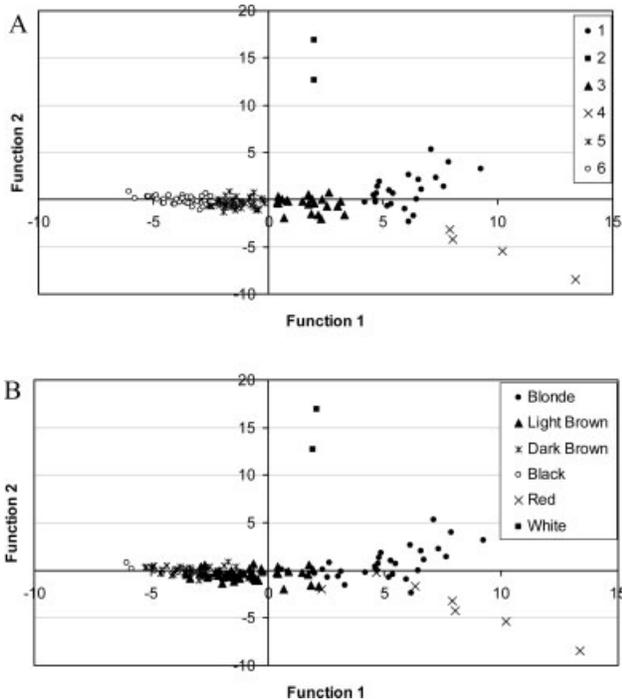


Fig. 4. The first two functions used to discriminate the six clusters as determined by log-likelihood (Analysis 2) were applied to the $L^*a^*b^*$ data and grouped by Cluster (A) and by observer-reported color (B) where $F1 = -0.013 + 0.073L^* + 0.891a^* + 0.573b^*$ and $F2 = -8.477 + 0.338L^* - 0.932a^* - 0.224b^*$.

gorize hair color for quantitative research. Although hair color can be perceived as a continuum (as seen in Fig. 1), it can be divided into categories or clusters based on spectrophotometric measurements in the CIE $L^*a^*b^*$ color space and these clusters can be well discriminated from each other. The discrepancies between self-reported hair color and observer-reported hair color and the poor separation of observer-reported color groups compared to those defined by clustering analysis emphasize the importance for objective measurement of individuals to be included in research studies. For repeatability and validity of studies, phenotypes must be defined as well as possible and the objective measurement methods and analysis strategies presented here may be of assistance in this regard.

The discriminant functions used to make classifications show some interesting results. It can be seen that, when using the CIE $L^*a^*b^*$ color space, the b^* component (yellow) is the most important in describing the variation and grouping of hair colors, followed closely by L^* , in the clusters examined in this study. This suggests that using one component is not sufficient for examining hair colors, as has been done in previous studies (Shriver and Parra, 2000) and that yellow (b^*) is the primary contributor instead of L^* . This, interestingly, corresponds to the 2004 study by Naysmith et al. (2004) where b^* was shown to have the strongest relationship to MC1R genotype. The red component of hair contributes very little information to discriminating groups. The reduction of three color components to one or two discriminant functions makes the examination and classification of individuals easier to represent.

It is also interesting to note that, while not exact, clustered groupings follow the same pattern that self- and observer-reported colors do. How someone decides to report a hair color may be an artifact of the arbitrary naming of colors or of personal bias or it may also be due to human perception of color, which corresponds to the CIE $L^*a^*b^*$ system of measurement. The human eye is a complicated structure with retinal receptors (rods and cones) that are sensitive to changes in lightness and color (Hunt, 1998). Very generally, when receiving a visual signal, these receptors judge the brightness of that signal (mainly the rods) as well as analyzing the hue of a signal as green or red and as blue or yellow (mainly the three types of cones), with this combination of signals being perceived as a color (Hunt, 1998). This system of human trichromatic vision may be why similar groups are defined by both the human reporting of color and by the spectrophotometry and clustering method described here. The lack of sensitivity in biological perception may contribute to the variability found. As mentioned previously, one unit on the $L^*a^*b^*$ scale is defined as the limit of human discrimination between colors (TASI, 2004). The difference between members of different clusters can be less than this in one or more of the three color axes and may contribute to the inconsistency between reported color groups and clusters determined by spectrophotometric data and clustering algorithms.

The particular clustering algorithms and program options used will depend on the intended use of the data. For example, a medical genetic association study may require a Fair/Dark division, where the prediction of hair color in a forensic study may require more specific groupings. This study has illustrated the possibilities by using a two-step cluster analysis in SPSS; however, there are many other programs and clustering algorithms available. Having objectively measured colors allows researchers to group individuals in clearly defined ways and to change these groups as the study demands without subjectivity or collecting additional data.

For practical purposes, it is acknowledged that the exact functions and cluster details may change as larger samples are analyzed and that reflective spectrophotometry, while accurate, may be inconvenient for measurements on very large numbers of people or for investigators and the authors hope to address these issues in the near future.

ACKNOWLEDGMENTS

The authors thank Dr. Neil Barnett at Victoria University for his assistance with the statistical analysis as well as all the participants in this study.

LITERATURE CITED

- Ancans J, Flanagan N, Hoogduijn MJ, Thody AJ. 2003. P-locus is a target for the melanogenic effects of MC1R signalling: a possible control point for facultative pigmentation. *Ann NY Acad Sci* 994:373-377.
- Aoki K. 2002. Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Ann Hum Biol* 29: 589-608.
- Borges CR, Roberts JC, Wilkins DG, Rollins DE. 2001. Relationship of melanin degradation products to actual melanin content: application to human hair. *Anal Biochem* 290:116-125.
- Box NF, Wyeth JR, O'Gorman LE, Martin NG, Sturm RA. 1997. Characterization of melanocyte stimulating hormone receptor

- variant alleles in twins with red hair. *Hum Mol Genet* 6:1891–1897.
- Duffy DL, Box NF, Chen W, Palmer JS, Montgomery GW, James MR, Hayward NK, Martin NG, Sturm RA. 2004. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum Mol Genet* 13:477–461.
- Ford A, Roberts A. 1998. Colour space conversions. www.poynton.com/PDFs/coloureq.pdf
- Frost P. 2007. Human skin-color sexual dimorphism: a test of the sexual selection hypothesis. *Am J Phys Anthropol* 133: 779–780.
- Grimes EA, Noake PJ, Dixon L, Urquhart A. 2001. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci Int* 122:124–129.
- Ha T, Naysmith L, Waterson K, Oh C, Weller R, Rees JL. 2003. Defining the quantitative contribution of the melanocortin 1 receptor (MC1R) to variation in pigmentary phenotype. *Ann NY Acad Sci* 994:339–347.
- Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL. 2000. Evidence for variable selective pressures at MC1R. *Am J Hum Genet* 66:1351–1361.
- Hunt RWG. 1998. *Measuring colour*. Kingston-upon-Thames, 3rd ed. England: Fountain Press. p 19–30.
- Krause J, Orlando L, Serre D, Viola B, Prüfer K, Richards MP, Hublin JJ, Hänni C, Derevianko AP, Pääbo S. 2007. Neanderthals in Central Asia and Siberia. *Nature* 449:902–904.
- Lalueza-Fox C, Rompler H, Caramelli D, Staubert C, Catalano G, Hughes D, Rohland N, Pilli E, Longo L, Condemi S, de la Rasilla M, Fortea J, Rosas A, Stoneking M, Schoneber T, Bertranpetit J, Hofreiter M. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318:1453–1455.
- Madrigal L, Kelly W. 2007a. Human skin-color sexual dimorphism: a test of the sexual selection hypothesis. *Am J Phys Anthropol* 132:470–482.
- Madrigal L, Kelly W. 2007b. Human skin-color sexual dimorphism: a test of the sexual selection hypothesis. Reply to Frost (2007). *Am J Phys Anthropol* 133:780–781.
- Naysmith L, Waterson K, Ha T, Flanagan N, Bisset Y, Ray A, Wakamatsu K, Ito S, Rees JL. 2004. Quantitative measures of the effect of the melanocortin 1 receptor on human pigmentary status. *J Invest Dermatol* 122:423–428.
- Nogueira ACS, Joekes I. 2004. Hair color changes and protein damage caused by ultraviolet radiation. *J Photochem Photobiol B* 74:109–117.
- Parra EJ. 2007. Human pigmentation variation: evolution, genetic basis, and implications for public health. *Yearbk Phys Anthropol* 50:85–105.
- Parra EJ, Kittles RA, Shriver MD. 2004. Implications of correlations between skin color and genetic ancestry for biomedical research. *Nat Genet* 36:s54–s60.
- Rees JL. 2003. Genetics of hair and skin color. *Annu Rev Genet* 37:67–90.
- Shriver MD, Parra EJ. 2000. Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. *Am J Phys Anthropol* 112:17–27.
- Slominski A, Tobin DJ, Shibahara S, Wortsman J. 2004. Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiol Rev* 84:1155–1228.
- Sturm RA, Duffy DL, Box NF, Newton RA, Shepherd AG, Chen W, Marks LH, Leonard JH, Martin NG. 2003. Genetic association and cellular function of MC1R variant alleles in human pigmentation. *Ann NY Acad Sci* 994:348–358.
- Takahashi T, Nakamura K. 2004. A study of the photolightening mechanism of blond hair with visible and ultraviolet light. *J Cosmet Sci* 55:291–305.
- Takahashi T, Nakamura K. 2005. A study of the photolightening mechanism of red hair with visible and ultraviolet light: comparison with blond hair. *J Cosmet Sci* 56:47–56.
- TASI. 2004. *Colour theory: understanding and modelling colour*. Technical Advisory Service for Images. Bristol, UK: University of Bristol.
- Westerhof W. 2006. The discovery of the human melanocyte. *Pigment Cell Res* 19:183–193.



A comparison of hair colour measurement by digital image analysis with reflective spectrophotometry

Michelle R. Vaughn^{a,*}, Roland A.H. van Oorschot^b, Swati Bainsur-Hudson^c

^aVictoria University, School of Molecular Sciences, Hoppers Lane, Melbourne, Werribee, VIC 3030, Australia

^bVictorian Police Forensic Services Department, Macleod, VIC, Australia

^cVictoria University, School of Biomedical and Health Sciences, Melbourne, VIC, Australia

ARTICLE INFO

Article history:

Received 16 May 2008

Received in revised form 5 November 2008

Accepted 7 November 2008

Available online 16 December 2008

Keywords:

Hair colour

Digital image analysis

Reflective spectrophotometry

Forensic science

Genetics

Macroscopic

Investigative tool

ABSTRACT

While reflective spectrophotometry is an established method for measuring macroscopic hair colour, it can be cumbersome to use on a large number of individuals and not all reflective spectrophotometry instruments are easily portable. This study investigates the use of digital photographs to measure hair colour and compares its use to reflective spectrophotometry. An understanding of the accuracy of colour determination by these methods is of relevance when undertaking specific investigations, such as those on the genetics of hair colour. Measurements of hair colour may also be of assistance in cases where a photograph is the only evidence of hair colour available (e.g. surveillance). Using the CIE $L^*a^*b^*$ colour space, the hair colour of 134 individuals of European ancestry was measured by both reflective spectrophotometry and by digital image analysis (in V++). A moderate correlation was found along all three colour axes, with Pearson correlation coefficients of 0.625, 0.593 and 0.513 for L^* , a^* and b^* respectively (p -values = 0.000), with means being significantly overestimated by digital image analysis for all three colour components (by an average of 33.42, 3.38 and 8.00 for L^* , a^* and b^* respectively). When using digital image data to group individuals into clusters previously determined by reflective spectrophotometric analysis using a discriminant analysis, individuals were classified into the correct clusters 85.8% of the time when there were two clusters. The percentage of cases correctly classified decreases as the number of clusters increases. It is concluded that, although more convenient, hair colour measurement from digital images has limited use in situations requiring accurate and consistent measurements.

© 2008 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The objective measurement of hair colour has been a topic of several studies in the past few years and is an important part of future research in forensic science [1–4]. Having objectively defined hair colours and the ability to classify them, according to the requirements of a particular study, would allow for increased consistency between research studies and would also allow for greater agreement in discussion between individuals such as investigators, witnesses and forensic scientists. In addition, hair colour is a physical feature that genetic researchers are interested in as part of a ‘phenotype profile’; a way to predict what someone looks like from DNA for investigative purposes [5]. This also requires standardised measurement of hair colour [5]. Reflective spectrophotometry has often been used to measure hair colour and is an established

method [1–6]. A previous study by the authors [7] has shown that hair colours measured by reflective spectrophotometry can be classified into meaningful groups, with those groups depending on the intended use of the data. For example, a medical genetics case–control study may be interested in a light/dark separation of colours, where those interested in forensic prediction may be more interested in separating colours into smaller, more specific groups.

While reflective spectrophotometry is an established method for measuring hair colour at the macroscopic level, it can be cumbersome to use on a large number of individuals and not all reflective spectrophotometry instruments are easily portable. This study investigates the use of digital photographs to measure hair colour and compares its use to reflective spectrophotometry. An understanding of the accuracy of the colour determination by these methods is of relevance when undertaking specific investigations, such as those on the genetics of hair colour. Measurement of hair colour by digital image analysis may also be of assistance in cases where a photograph is the only evidence of hair colour available (e.g. surveillance).

* Corresponding author. Tel.: +61 4 2580 5309; fax: +61 3 9919 8284.

E-mail address: michelle.vaughn@live.vu.edu.au (M.R. Vaughn).

Digital image analysis has previously been used to measure hair colour in forensic science; however it has generally been used to measure the colour during the microscopic examination of single hairs and not the macroscopic colour of hair on the head that this study is concerned with [8]. This has been done to assist the traditional forensic hair examination, with the purpose of attributing a hair to a specific donor. Digital images have been used extensively to measure the colour of many other objects, such as vegetables, soil and chocolate [9–11], and have been used to measure skin colour [12–14]. Herbin et al. found that the colour of skin lesions measured from 12 digitised images of colour slides was highly correlated with values obtained by reflective spectrophotometry, with correlation coefficients of 0.97, 0.99 and 0.98 for L^* , a^* and b^* of the CIE $L^*a^*b^*$ model (see below) respectively. Takiwaki et al. [13] achieved similarly good results with eight videomicroscopic images of skin.

To objectively measure hair colour, a method of measuring colour must first be chosen. There are numerous models or 'colour spaces' used to describe colour. The CIE $L^*a^*b^*$ model, developed by the Commission Internationale de l'Eclairage, measures colour on three axes that correspond to trichromatic human perception and reflect the degree of change in colour that humans can perceive [15]. This model has the benefit of corresponding to the human perception of colour and has the additional benefit giving a grid point for each specific colour.

The lightness, or intensity, of a colour is measured on the ' L^* ' axis on a scale from 0 (black) to 100 (white). The colour is also measured on the ' a^* ' axis that gives a value from -100 (green) to $+100$ (red). The ' b^* ' axis measures colour from -100 (blue) and $+100$ (yellow). One unit on the L^* , a^* or b^* axes is considered to be the smallest difference the human eye can perceive [16]. This grid point allows for the mathematical comparison of colours and also allows colours to be corrected for differing conditions. It should be noted that, theoretically, the a^* and b^* axes have no maximum or minimum values. Our research has used the cut off point of ± 100 because this is the practical limit of the software and instrument used to measure colour [17].

When measuring colour from digital images, it is important to note that digital images are most often displayed and measured in the Red-Green-Blue (RGB) colour space of the computer monitor. There are various programs that may be used to measure this colour and then to convert it into the colour space required (CIE $L^*a^*b^*$ in this case) using standard mathematical algorithms [12]. This study considers two such programs.

The aim of this study is to compare the results obtained from measuring the colour from digital images (DI) of hair colour (at the macroscopic level) to the measurements of the same individuals using reflective spectrophotometry (RS) as a standard, to see how the measurements correspond. This will help determine to what extent digital image analysis can be used instead of reflective spectrophotometry and, if so, in what situations.

2. Materials and methods

2.1. Subjects

Subject recruitment and sampling procedures were conducted with the approval of the Victoria University Human Research Ethics Committee (approval number HRETH06/156). In total, 134 subjects were included in this analysis. All volunteers recruited for this study had their natural hair colour at the time of sampling. Most volunteers were between the ages of 18 and 35 and were of European ancestry (132 subjects). The exceptions to these criteria include two older individuals with white hair. The subject criteria were chosen to reinforce the study aims of examining natural variation in adult, European hair colour, with white haired individuals included for comparison.

2.2. Sampling

After obtaining written consent, subjects filled out a questionnaire that, in addition to confirming their age, ancestry and natural hair colour status, asked them to report how they saw their own hair colour. Hair colour was reported to be one of;

Black, Dark Brown, Light Brown, Blonde, Red or White. Colour assessment by an observer (the researcher taking the samples) was also recorded.

Hair colour was measured by reflective spectrophotometry (RS) using a Minolta CR-300 Chroma Meter (Konica Minolta, North Ryde NSW, Australia). The CR-300 has an 8 mm measurement area and uses D/0 viewing geometry. The instrument was calibrated using a white tile and a light source input setting of 'D65' which represents daylight without spectral highlights. The instrument was set to measure in the CIE $L^*a^*b^*$ format. Hair colour was measured on the left, right and back of the head and was measured five times at each spot, with the instrument in contact with the head. The mean of the 15 measurements is the focus of subsequent statistical considerations.

Digital images were taken of the left, right and back of the head with a CASIO 4.0 mega pixel camera (model QV-R40) with an image size of 640×480 pixels and with picture quality set to 'fine'. On 'collection' mode (for taking close-up photographs), the flash was turned off, the white balance tool was set to 'sunlight' and all other settings were left at the default. Each time a set of images was to be taken under a different light source (on a different day, different time of day, different room, etc.) a piece of new white paper was photographed in the same position as the head of the subject (three sides) and measured using the Minolta CR-300. This 'white image' was later used to correct colour for light source. All images were taken at a distance of 15 cm.

2.3. Digital image preparation and analysis

Software packages used included:

- JASC Paint Shop Photo Album 4.0 by Jasc Inc.
- Adobe Photoshop Version 9.0.2[®] 1990–2005 Adobe Systems Inc.
- IrfanView 3.92[®] 1996–2004 Irfan Skijan; Vienna University of Technology.
- MATLAB 4.0[®] 1994–2006 MathWorks with the Image Processing Toolbox and modules written to convert RGB colour measurements to $L^*a^*b^*$ measurements [9].
- V++ Version 4.0[®] 1990–2002 Digital Optics Ltd. New Zealand, with modules written to convert RGB colour measurements to $L^*a^*b^*$ measurements [9,18].
- ImageJ[®] 1997–2006 by W.S. Rasband at the US National Institutes of Health, with the RGB-to-Lab conversion plugin installed [19].

Each digital image of the hair was cropped to 100×100 pixels in the area closest to the centre of the image (where the RS measurements were taken) with JASC Photo Album, Adobe Photoshop or IrfanView as available. Care was taken to avoid obvious shadows or spectral highlights that result from the curvature of the skull.

Initially, images were analysed using MATLAB to convert the RGB colour measurements to $L^*a^*b^*$ and to determine the mean colour values for each image; however, this program is rather difficult and tedious for a new user. The RGB to $L^*a^*b^*$ conversion modules written for MATLAB were instead used within V++ [18], a more user-friendly program, which results in identical values as those determined by MATLAB (results not shown). Therefore, V++ was used to analyse all subsequent images.

For each image, the L^* , a^* and b^* values were calculated. Images of the same colour taken under different light sources will appear different [12], so an effort was made to correct this colour in a manner similar to a previous study using the RGB colour space [12], with the understanding that perfect correction is unlikely and that the greatest difficulty in correcting the colours (a^* and b^* axes) will be found when the lightness of the image differs. The new white paper, mentioned above, was measured with the Minolta Chroma Meter CR-300 and found to have L^* , a^* and b^* values of 93.4, 1.16 and -2.06 respectively. Using this as a reference point, L^* , a^* and b^* colour points were added or subtracted depending on how much the white reference paper deviated from its Minolta CR-300 determined values under the various sampling situations. For example, if L_i is the L value of the colour image, L_w is the L value of the white image and L_c is the L value of the corrected image, then: $L_c = (93.4 - L_w) + L_i$.

2.4. Correction

To examine the accuracy of the correction formula, images were taken of white, pink and blue paper under four different light sources (i) fluorescent light, (ii) fluorescent light and natural sunlight, (iii) incandescent light and natural sunlight and (iv) bright natural sunlight. The images of the paper were cropped to 100×100 pixels and analysed for colour with V++. The pink and blue colours were corrected using the white values obtained from V++, using the above formula and compared to the RS measurements (true colour) of the paper.

2.5. Program comparison

To examine whether another program is better able to match the RS values through its algorithm or data handling, the MATLAB/V++ module for RGB to $L^*a^*b^*$ colour conversion was compared to that of the ImageJ program. Values were calculated for the pink, blue and white paper images and corrected using the white paper values calculated from ImageJ and with the above formula. The corrected ImageJ values were compared to the RS measurements and to the V++ values.

2.6. Statistics

SPSS 15.0 for Windows (©SPSS Inc. 2006, Chicago, IL) was used to analyse the colour data. Basic statistical tests were performed, including a One-Sample *T*-test (two tailed) for a difference in mean, a Paired *T*-test (two tailed) and an *F*-test for a difference in variance [20].

For details on the cluster analysis using reflective spectrophotometric data, please see Vaughn et al. [7]. Generally, cluster analysis seeks to identify natural subgroups within a population by minimising within-group variation while maximising between-group variation. To characterise and evaluate the clusters a separate, discriminant analysis was then performed. Discriminant analysis commences from the perspective of there being a known number of population sub-groups. From the population, there are a number of individuals whose sub-group classification is known. The data from these individuals is analysed in an attempt to build a profile of sub-group membership and subsequently use this profile to classify new individuals (whose sub-group membership is otherwise unknown). After the analysis, SPSS determines which variables are most important in discriminating the groups, and how good this model will be at predicting group membership of future individuals. It determines the predictive value of the model by removing each individual separately, reanalysing the data and then predicting the membership of that case. The percent correctly classified is reported.

3. Results

3.1. Colour correction

When comparing the colour mean values of pink paper measured by digital image analysis to the reflective spectrophotometric value, it was found that the DI a^* component values have means that are significantly lower than the RS measurements, whether the correction formula is used or not. Using a One sample *T*-test, there was a 95% confidence interval of -1.51 to -5.96 for the difference between the uncorrected DI measurement for a^* and the RS measurement and from -0.89 to -8.27 for the difference between the corrected DI measurement for a^* and the RS measurement, thus neither of these CI ranges included 0 (null hypothesis = 0 difference). There was no significant difference between the RS and DI means for the L^* or b^* values. For the blue paper, only the corrected L^* value was significantly different from the RS measurements with a 95% confidence interval (One sample *T*-test) of 3.78 – 9.90 .

Although not all of the means were significantly different, it was noted that for the L^* and b^* values of Pink and Blue paper (four out of six measurements) the confidence intervals for corrected digital measurements were narrower than those for the uncorrected formulas, and so an *F*-test for the equality of variance was conducted (between the uncorrected and corrected values for pink and blue paper). The results indicate that the variance for L^* of pink paper and the L^* and b^* components of blue paper are significantly smaller using the correction formulas (*F*-test for variance, p -value < 0.05). It was decided to use the correction formula for further experiments due to the closer means (though not significant) and the significantly smaller variances.

3.2. Program comparison

To test whether the colours obtained by the programs V++ and ImageJ are any different, a Paired *T*-test of the means was used for the corrected L^* , a^* and b^* values from each program. The only significant difference (p -value = 0.002) was that the a^* value for the blue paper was slightly higher in ImageJ (-5.19) than in V++ (-5.99).

The variance of the measurements as determined by the two programs was also compared. There was no significant difference (*F*-test for variance; p -values > 0.1) between the variance of colour measurements as determined by V++ and ImageJ. It can be concluded that the L^* , a^* , b^* colour measurement functions of ImageJ and V++ can be used interchangeably.

3.3. Intra-individual comparison

To test the consistency of the digital image colour measurements, 10 sets of digital images were taken of one individual (at the right, left and back of the head). The mean of the L^* , a^* and b^* values for each set (of three images) was calculated and standard deviations of 2.43, 0.42, 0.74 were obtained across the 10 sets for the L^* , a^* and b^* values, respectively. This illustrates the low amount of variation found across multiple measurements taken of one individual under the same conditions.

3.4. Digital image comparison to reflective spectrophotometry

After determining that the correction formulas produce more consistent results and that two programs selected to analyse the digital images are almost the same, the entire European sample (134 individuals) was analysed to determine how digital image (DI) colour analysis compares to the colours as determined by reflective spectrophotometry (RS).

Initially, the digital image colour measurements were compared directly to the RS measurements for the L^* , a^* and b^* axes. A correlation was found along all three axes, with Pearson correlation coefficients of 0.625, 0.593 and 0.513 for L^* , a^* and b^* respectively (p -values = 0.000). This is a significant correlation, but of only moderate strength. The L^* axis measurements are most highly correlated, followed by a^* , then by b^* .

To see if there is a pattern in the differences between DI and RS measurements, a Paired *T*-test for the difference between the means obtained by DI and RS was performed. Means are significantly overestimated by digital image analysis for all three colour components. The L^* value is significantly higher by an average of 33.42 units, the a^* value higher by an average of 3.38 units and the b^* value higher by an average of 8.00 units (p -values = 0.000).

To investigate whether the overestimation error was uniform across the sample, or if light or dark hair is more prone to overestimation, the difference between the DI and RS measurements was compared to the original colour (RS measurement). The differences in L^* and b^* showed no significant correlation, indicating that the amount of overestimation is not related to the original hair colour for these components; however, there was a low amount of correlation (Pearson's Coefficient of 0.326) between the overestimation of the DI a^* and the RS a^* values (p -value = 0.000), indicating that there is a slight tendency for the difference in a^* values to become greater as hair becomes more red.

As the separate colour components (L^* , a^* and b^*) measured by RS and DI do not correspond well with each other, we investigated whether when combined they would place individuals into the same colour categories as those determined by RS measurements.

In a previous study, we showed that the RS measurements can be used to break this same sample set into defined groups using a Two-Step Cluster analysis [7]. These groups were then evaluated using a discriminant analysis to classify cases into the correct groups. In that study, the best results were obtained for three analyses: (1) two clusters using a log-likelihood probability method; (2) six clusters using a log-likelihood method and (3) four clusters using a Euclidian distance method. Each analysis obtained classification scores above 95% [7]. This was better than the 73.1% discrimination obtained by using observer-reported clusters [7].

For the present study, each individual was assigned to a cluster based on their RS data (for each of the three analyses). These clusters were then re-evaluated, using their DI data and the percent of cases correctly classified is reported in Table 1. Table 1 indicates that as the number of clusters increases, the percent of individuals correctly classified decreases. Table 1 also shows that

Table 1

Percent of cases correctly classified by a discriminant analysis on Clusters determined by RS or the observer reported colour group (ORC), using DI data.

Analysis	Number of clusters	Percent correctly classified by digital image analysis
1	2	85.8%
2	6	61.2%
3	4	76.9%
ORC	6	51.5%

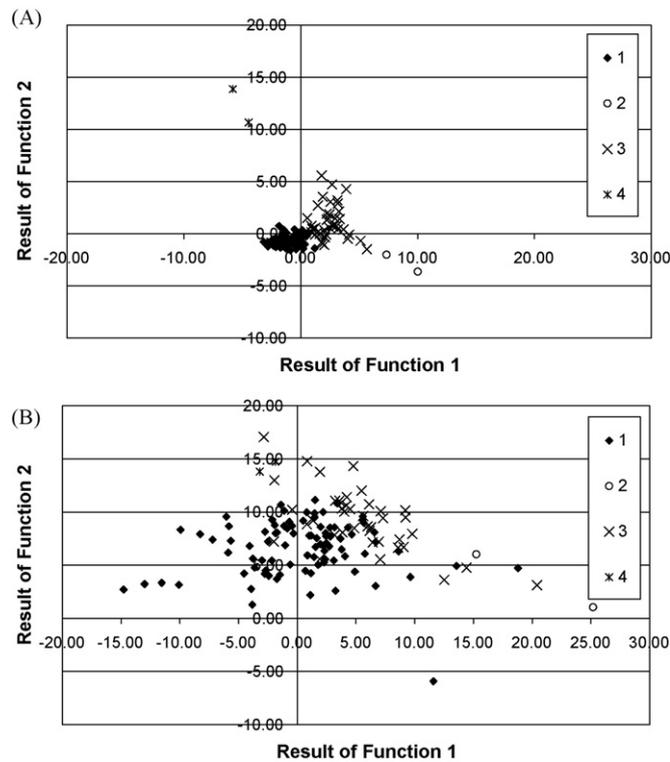


Fig. 1. RS clusters with RS data (A) compared to DI data (using RS discriminant functions) (B) for Analysis 3. Clusters from Analyses 3 [7] are numbered 1–4.

the percent of cases correctly classified in the six observer-reported colour groups (51.5%) was lower than for any of the RS analyses.

Fig. 1 illustrates an example of the use of discriminant functions in classifying individuals (using the results from Analysis 3). The discriminant functions determined by RS data ($F1 = 0.107 - 0.122L^* + 0.669a^* + 0.486b^*$ and $F2 = -8.894 + 0.297L^* - 0.442a^* - 0.117b^*$) were used to plot individuals onto a grid using first their RS measurements (Fig. 1A) and then their DI measurements (Fig. 1B). Fig. 1A shows the good separation of the four clusters determined by Analysis 3 while Fig. 1B shows the groups merging into one another.

4. Discussion

The presented data have shown that although measuring colour from digital images is convenient, the colours measured and classified do not correspond very well to those colours measured by reflective spectrophotometry and would have limited use in studies requiring accurate measurement of hair colour, such as genetic association studies. They also do not provide an accurate description of hair colour for crime investigators to utilise as visual identifiers or in description announcements. Some use might be found in distinguishing between a small number of clusters with digital images, as in the two clusters of Analysis 1 (which showed

85.8% of cases correctly classified); however, as the number of clusters or groups increases, there is a lower correspondence. Observer-reported colours were not well discriminated by RS data (73.1% of cases correctly classified) [7] and were more poorly discriminated by using digital image analysis (51.5% of cases correctly classified). These results support the assertion that objectively defined and measured colours, rather than observer or self-reported colours are necessary to classify individuals for further study and at this time, this is best achieved through reflective spectrophotometric measurement.

It should be noted that the images in this study were all taken with the same camera under the specific conditions of a well-lit room and from a close, specified and identical distance and all treated in the same way. Multiple analysis of one individual showed very low variation in the results. Any additional variability in conditions would contribute to less similarity between the digital image analysis and a reflective spectrophotometry standard. If variability could be reduced, through improvement in photographic conditions or through more accurate correction for varying conditions, results may improve. The act of photographing an object introduces error in colour reproduction due to the nature of the camera and image production and is an approximation of the visual field (whether digital or with film) [21]. Different models of camera may have different abilities to reproduce colour accurately and may improve results. The process of measuring RGB values and converting RGB measurements to CIE $L^*a^*b^*$ colour space may also result in the loss of some colour information.

It may also be that there are components of hair colour that are omitted when measuring colour directly from a digital photograph. For example, features such as hair texture, thickness, luster or the natural inhomogeneity of hair colours may have a different effect on the analysis of colour from a digital image than they have on reflective spectrophotometry or on human perception. With these things in mind, the improvement of digital cameras and further research into the digital photography of hair would assist in developing methods that provide a more accurate measurement and classification of hair colour that would be of use to researchers, investigators and forensic scientists.

Acknowledgements

We would like to thank Bruce Comber (Australian Federal Police) and Dr. Alasdair McAndrew (Victoria University) for their assistance with digital image analysis software, as well as Dr. Neil Barnett (Victoria University) for his assistance with the statistical analysis. We would also like to thank all the participants in this study. This research was supported by an Endeavour International Postgraduate Research Scholarship (Australia) and a National Science and Engineering Research Council Postgraduate Scholarship (Canada).

References

- [1] M.D. Shriver, E.J. Parra, Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry, *Am. J. Phys. Anthropol.* 112 (2000) 17–27.
- [2] E.J. Parra, R.A. Kittles, M.D. Shriver, Implications of correlations between skin color and genetic ancestry for biomedical research, *Nat. Genet.* 36 (2004) s54–s60.
- [3] L. Naysmith, K. Waterson, T. Ha, N. Flanagan, Y. Bisset, A. Ray, K. Wakamatsu, S. Ito, J.L. Rees, Quantitative measures of the effect of the melanocortin 1 receptor on human pigmentation status, *J. Invest. Dermatol.* 122 (2004) 423–428.
- [4] A.C.S. Nogueira, I. Joekes, Hair color changes and protein damage caused by ultraviolet radiation, *J. Photochem. Photobiol. B* 74 (2004) 109–117.
- [5] T.N. Frudakis, *Molecular Photochemistry*, Elsevier Inc., Burlington, MA, 2008.
- [6] S.N. Shekar, D.L. Duffy, T.N. Frudakis, G.W. Montgomery, M.R. James, R.A. Sturm, N.G. Martin, Spectrophotometric methods for quantifying pigmentation in human hair—influence of MC1R genotype and environment, *Photochem. Photobiol.* 84 (2008) 719–726.
- [7] M. Vaughn, R. van Oorschot, S. Baidur-Hudson, Hair color measurement and variation, *Am. J. Phys. Anthropol.* 137 (2008) 91–96.

- [8] J. Bednarek, An attempt to establish objective criteria for morphological examinations of hairs using the image analysis system, *Probl. Forensic Sci. LVI* (2004) 65–77.
- [9] Z. Shen, Colour differentiation in digital images (MSc Thesis), School of Computer Science and Mathematics, Victoria University, Melbourne, 2003, p. 167.
- [10] N. Levin, E. Ben-Dor, A. Singer, A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments, *Int. J. Remote Sens.* 26 (2005) 5475–5492.
- [11] B. Kumara, S. Jinap, Y.B. Che Man, M.S.A. Yusoff, Note: comparison of colour techniques to measure chocolate fat bloom, *Food Sci. Technol. Int.* 9 (2003) 295–299.
- [12] M. Herbin, A. Venot, J. Y. Devaux, C. Piette, Color quantitation through image processing in dermatology, *IEEE Trans. Med. Imag.*, 9 (1990) 262–269.
- [13] H. Takiwaki, H. Miyamoto, K. Ahsan, A simple method to estimate CIE- $L^*a^*b^*$ values of the skin from its videomicroscopic image, *Skin Res. Technol.* 3 (1997) 42–44.
- [14] G.E. Pierard, Eemco guidance for the assessment of skin colour, *J. Euro. Acad. Dermatol. Venereol.* 10 (1998) 1–11.
- [15] A. Ford, A. Roberts, *Colour Space Conversions* (1998).
- [16] TASI, Colour theory: understanding and modelling colour, 2004, <http://www.tasi.ac.uk/advice/creating/pdf/colour.pdf> accessed on December 5, 2007.
- [17] S. Napier, Personal communication, Biolab Group Australia, 2007.
- [18] B. Comber, Personal communication, Australian Federal Police, 2007.
- [19] W.S. Rasband, ImageJ, US National Institutes of Health, Bethesda, Maryland, USA, 1997–2006, <http://rsb.info.nih.gov/ij/>.
- [20] J.T. McLave, T. Sincich, *Statistics*, Prentice-Hall Inc., Upper Saddle River, NJ, 2003.
- [21] R.E. Jacobson, S.F. Ray, G.G. Attridge, N.R. Axford, *The Manual of Photography: Photographic and Digital Imaging*, Focal Press: Elsevier Science Inc., San Diego, USA, 2000.

A Comparison of Macroscopic and Microscopic Hair Color Measurements and a Quantification of the Relationship between Hair Color and Thickness

Michelle R. Vaughn,^{1,*} Elizabeth Brooks,² Roland A.H. van Oorschot,³ and Swati Baidur-Hudson⁴

¹School of Molecular Sciences, Victoria University, Melbourne, VIC, Australia

²Forensic and Technical, Australian Federal Police, Canberra, ACT, Australia

³Victorian Police Forensic Services Department, Macleod, VIC, Australia

⁴School of Biomedical and Health Sciences, Victoria University, Melbourne, VIC, Australia

Abstract: Microscopic images of hair from 36 individuals were evaluated for color (using the CIE $L^*a^*b^*$ color space), hair thickness, and the pattern of pigmentation. These measurements were compared to each other and to reflective spectrophotometric measurements of macroscopic hair color. Three areas on three strands of hair were examined for each individual. Despite a large ΔE^* value, there is a strong relationship between macroscopic and microscopic L^* values that would be useful if one was trying to predict the color of a head of hair (macroscopic) from a strand (microscopic). A relationship between hair lightness (L^*) and thickness is also found at both the macroscopic and microscopic levels. The information acquired in this study may be useful to those conducting a microscopic examination of hair, especially in forensic science, as well as those researchers investigating macroscopic hair color and hair morphology for the purposes of forensic science, anthropology, and medical science. Further investigation into objectively measuring and quantifying the microscopic color and properties of hair will assist in increasing its validity as evidence in forensic casework.

Key words: forensic, hair, hair color, microscopic hair, macroscopic hair, hair thickness

INTRODUCTION

Hair is commonly encountered as physical evidence at crime scenes (Saferstein, 2001). Traditional forensic hair examination involves examining the specimen under a microscope and observing characteristics such as color, length, and diameter of the hair, as well as the distribution of pigment granules and other morphological features of the hair that may assist in determining the source of the hair (Saferstein, 2001; Bednarek, 2004; SWGMAT, 2005).

In an attempt to assist with the classification of hair colors in a forensic examination, Bednarek (2004) illustrated one way of measuring hair color at the microscopic level—calculating the red-green-blue (RGB) values of a digital image of the hair. The average RGB values for blond and brown hair were determined. Recent research has also been conducted on the measurement of hair color at the microscopic level for individualizing a hair (Brooks, 2007).

By using a discriminant analysis to assess color measurements, Brooks (2007) was able to correctly assign a strand of hair to a particular individual between 58 and 68% of the time (depending on what color model was used), based on known microscopic color measurements of the individuals' hair.

While accurate color measurements of hair at the microscopic level are of interest to forensic hair examiners, objectively measured hair color (at both the macroscopic and microscopic level) is also of interest to those investigating the genetic basis of physical features for use in predictive tests (Frudakis, 2008; Vaughn et al., 2008). Reflective spectrophotometry was used to measure the macroscopic hair color in this study because it is an established method and has been shown to be more consistent than digital image analysis (Vaughn et al., 2008, 2009).

The aim of this study was to determine the association between hair color measured at the microscopic level with color measured at the macroscopic level, utilizing the CIE $L^*a^*b^*$ color space. Because the macroscopic hair color is made up of thousands of strands of slightly different colors, it is important to know if the microscopic color of a strand

can indicate what the head of hair might look like and vice versa. This would be useful if a strand of hair is found as evidence and possibly also for the prediction of both macroscopic and microscopic color from genetic information. Examining morphological properties of the hair, such as thickness, and their associations with color provides further information on the physical basis of hair colors and may also be of relevance to those searching for genetic markers associated with hair color.

MATERIALS AND METHODS

In total, 36 individuals were included in this microscopic analysis. The study included 2 individuals with white hair, 3 individuals of non-European ancestry, and 31 individuals of European ancestry between the ages of 18–35, all with their natural hair color. These subjects were targeted to study natural, adult hair color variation as almost all natural hair color variation is found within individuals of European ancestry (Shriver & Parra, 2000). Each individual had his or her hair color measured by reflective spectrophotometry (RS) as described by Vaughn et al. (2008). In addition, several full-length strands of hair were cut from the scalp of these volunteers for microscopic analysis.

The CIE $L^*a^*b^*$ color space was used for both the macroscopic RS measurements and the microscopic digital image analysis. The CIE $L^*a^*b^*$ model, developed by the Commission Internationale de l'Éclairage, measures color on three axes that correspond to trichromatic human perception and reflect the degree of change in color that humans can perceive (Ford & Roberts, 1998). In this system, the lightness, or intensity, of a color is measured on the L^* axis on a scale from 0 (black) to 100 (white). The color is also measured on the a^* axis that gives a value from -100 (green) to $+100$ (red). The b^* axis measures color from -100 (blue) and $+100$ (yellow). One unit on the L^* , a^* , or b^* axes is considered to be the smallest difference the human eye can perceive (TASI, 2004). This grid point allows for the mathematical comparison of colors. It should be noted that, theoretically, the a^* and b^* axes have no maximum or minimum values. Our research has used the cut off point of ± 100 because this is the practical limit of the software and instrument used to measure color (Napier, 2007).

Hairs were mounted on Livingstone Premium glass slides (76.2×25.4 mm pathology grade) with Entellan mounting medium (Merck) and covered with Marienfeld cover slips (22×60 mm No. 1). The area to be examined was marked on the slide. Three random areas, at least 1 cm away from the hair root or tip, on each of three hairs were examined for each individual, and the mean of these measurements was used for most analyses (Fig. 1).

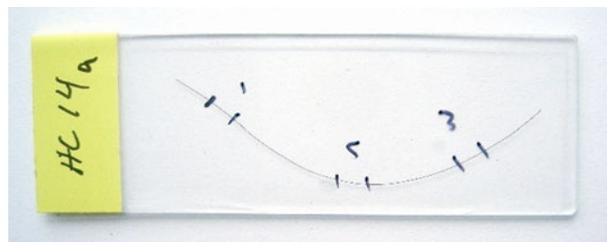


Figure 1. Hair prepared for microscopic examination.



Figure 2. Montaged image.

Equipment and Software Used to Generate Microscopic Images and to Measure Color

Equipment and methods are from Brooks (2007) with minor modification.

Leica DMLB is a microscope with a $50\times$ oil immersion lens ($500\times$ magnification; field of $276 \mu\text{m}$), using 10.5 V illumination (daylight equivalent) and with DLF (daylight filter) on.

Olympus DP70 is a 12.5 megapixel digital camera (Olympus Optical Co. Ltd. 2003), and the Image Management System consists of the DP Controller (Version 2.1.1.183) and DP Manager (Version 2.1.1.163). An image size of 1360×1024 pixels was used with ISO 200 sensitivity, exposure time set to “automatic,” 1% (of the field) contrast correction, and all images saved as TIF files.

Automontage-Pro 5.02.0096 software (by Syncrosopy, a division of Synoptics Group, UK) was used to montage images of the middle third of the hair (the area containing the greatest amount of pigmentation information) (Brooks, 2007). Hair thickness was estimated by the total depth of the optical field (Fig. 2).

V++ Version 4.0 (by Digital Optics Ltd. New Zealand, ©1990–2002) with a module written to convert RGB color measurements to $L^*a^*b^*$ measurements was used to extract



Figure 3. Extracted image of a light brown hair.



Figure 4. Extracted image of a blond hair.

the correct areas of the microscopic images and to measure the color of the extracted area (Figs. 3, 4).

RESULTS

Comparing Microscopic Color to Macroscopic Color

A difference in color can be measured by calculating ΔE^* , an Euclidian measure of the distance between two colors in the CIE $L^*a^*b^*$ space (Hunt, 1998):

$$\Delta E^* = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}.$$

The average ΔE^* between the macroscopic and microscopic color was 22.57 units, with a standard deviation of 5.41 units. However, this measurement does not give information about which components are different, and therefore each color axis was examined separately.

When comparing microscopic (Mc) color L^* , a^* , and b^* measurements directly to the macroscopic measurements determined by RS, regression analysis found that a^* and b^* measurements fail to show any significant linear relationship between Mc and RS (p -values = 0.317 and 0.288, respectively). RS and Mc L^* show a moderate, significant linear relationship with an $R = 0.631$ ($R^2 = 0.469$, p -value = 0.000). However, using curve estimation (multiple regression), it can be seen that there is a stronger quadratic relationship between the variables (Fig. 5).

Comparing Macroscopic Color to Hair Thickness

Macroscopic hair color, measured by RS, was compared to the estimates of hair thickness (the two white-haired individuals were not included). Macroscopic L^* (Fig. 6) and b^* were found to have a moderate, negative correlation with hair thickness, with correlation coefficients of -0.648 and -0.657 (p -values = 0.000), respectively. The a^* value showed a significant (p -value = 0.023) but weaker correlation of -0.390 . This means that thinner hair is correlated with being lighter and more yellow.

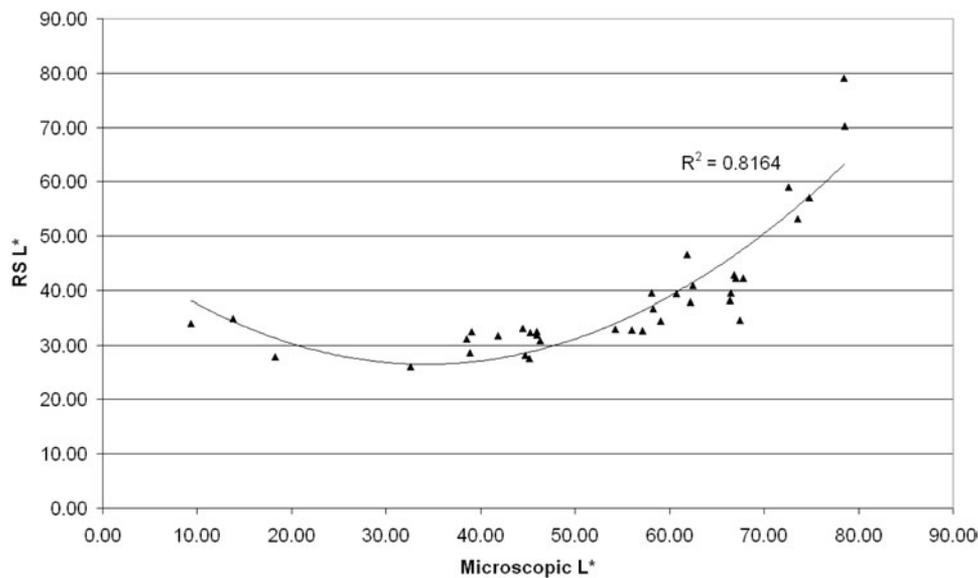


Figure 5. The quadratic relationship between microscopic L^* measurements and L^* measurements by RS. The trendline has the equation $y = 0.0188x^2 - 1.2854x + 48.505$ and $R^2 = 0.8164$.

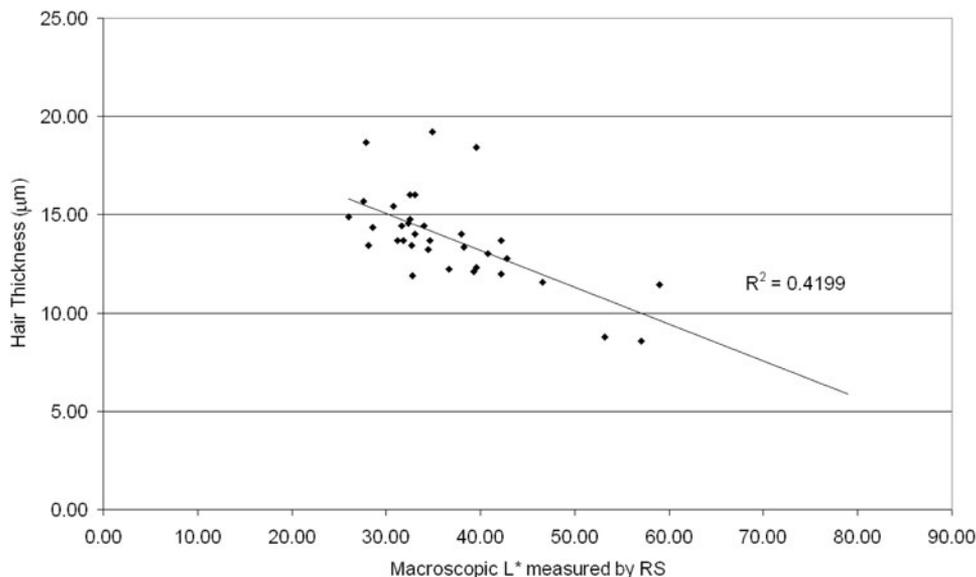


Figure 6. The relationship between the hair thickness (as measured by optical depth) and the macroscopic L^* value as measured by RS.

Comparing Microscopic Colors to Hair Thickness

Microscopic hair color was compared to the estimates of hair thickness. There were significant, negative correlations between the microscopic L^* values of hairs and their estimated thickness (in micrometers) with an $R = -0.705$ (p -value = 0.000) (Fig. 7). The a^* value was positively correlated to the hair thickness ($R = 0.567$, p -value 0.000), while b^* showed no significant relationship (data not shown).

DISCUSSION

This study illustrates the relationships between macroscopic and microscopic color and hair thickness. It illustrates that, despite the large ΔE^* value, there is a strong relationship between macroscopic and microscopic L^* values that would be useful if one were trying to predict the color of a head of hair (macroscopic) from a strand (microscopic) or if one

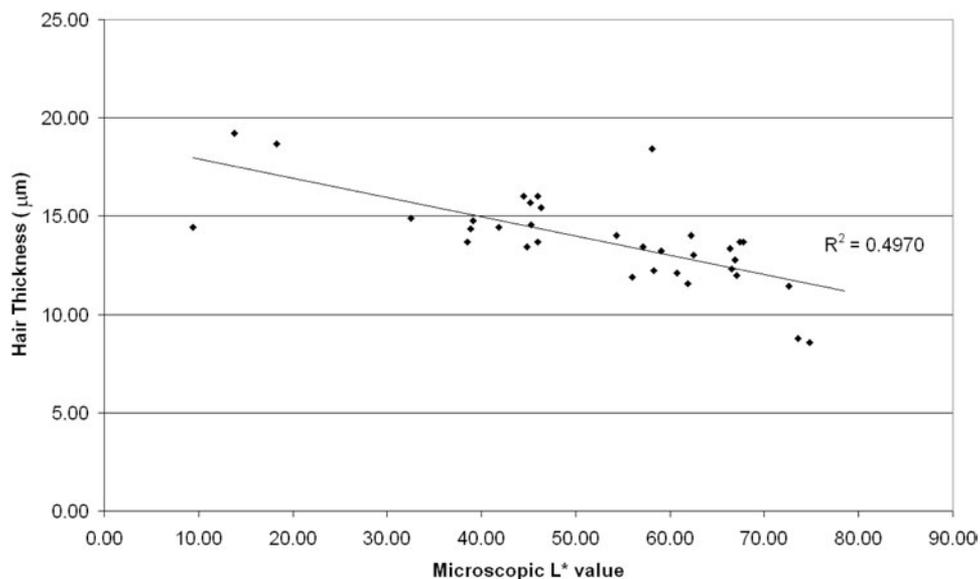


Figure 7. The relationship between the hair thickness (as measured by optical depth) and the microscopic L^* value.

was trying to predict hair color phenotype by genetic analysis and either the macroscopic or microscopic hair color was known to be more closely associated with a particular genetic variation. The relationship between hair lightness (L^*) and thickness is also clearly illustrated at both the macroscopic and microscopic levels.

While the reported relationships are useful and will assist in further research toward the objective measurement of hair colors, it can also be seen that there are weak or insignificant relationships between some aspects of the macroscopic and microscopic color, particularly when examining the a^* and b^* color components. These differences may be due to the fact that the number of hair strands that can be examined under the microscope were, and in the course of normal casework investigation most likely will be, a small proportion of the head of hair and larger sample sizes are difficult to obtain.

Color measurement by digital image analysis also presents some problems. In a previous study (Vaughn et al., 2009), the authors have shown that macroscopic hair color measured by digital images analysis does not correspond very well to the RS readings of the macroscopic hair color, and some of the same factors may be involved in measuring the color of microscopic images, such as the ability of the camera to reproduce color and the conversion from RGB to the $L^*a^*b^*$ color space. However, the high degree of consistency in conditions under a microscope should make microscopic digital images of hairs comparable to each other and therefore more consistent in their relationships with the macroscopic color.

The measurement of hair thickness could also be improved to increase the strength of relationships between thickness and color. Cross sections of the hair may be examined to determine the shape before measuring the thickness of hairs because using the optical depth may inadvertently use different sides of an oval-shaped hair for measurement and therefore be less precise.

The information acquired in this study may be useful to those conducting the microscopic examination of hair, especially in forensic science, as well as those researchers investigating macroscopic hair color and hair morphology for the purposes of forensic science, anthropology, and medical science. Further investigation into objectively measuring and quantifying the microscopic color and properties of hair will assist in increasing its validity as evidence in forensic casework.

ACKNOWLEDGMENTS

Our sincere thanks go to Australian Federal Police Forensic Services for the use of methods and facilities for conducting

this research. We would also like to thank Bruce Comber (Australian Federal Police) and Dr. Alasdair McAndrew (Victoria University) for their assistance with digital image analysis software. We would also like to thank all the participants in this study. This research was supported by an Endeavour International Postgraduate Research Scholarship (Australia) and a National Science and Engineering Research Council Postgraduate Scholarship (Canada).

REFERENCES

- BEDNAREK, J. (2004). An attempt to establish objective criteria for morphological examinations of hairs using the image analysis system. *Probl Forensic Sci* **LVI**, 65–77.
- BROOKS, E.M. (2007). An appraisal of the use of numerical features in the forensic examination of hair. University of Canberra M.Sc. Thesis. Available at <http://erl.canberra.edu.au/public/adt-AUC20080624.144159/index.html>.
- FORD, A. & ROBERTS, A. (1998). Colour space conversions. Available at www.poynton.com/PDFs/coloureq.pdf, accessed February 1, 2007.
- FRUDAKIS, T. N. (2008). *Molecular Photofitting*. Burlington, MA: Elsevier Inc.
- HUNT, R.W.G. (1998). *Measuring Colour*. Kingston-upon-Thames, UK: Fountain Press.
- NAPIER, S. (2007). Personal Communication, Biolab Group Australia.
- SAFERSTEIN, R. (2001). *Criminalistics*. Upper Saddle River, NJ: Prentice Hall.
- SHRIVER, M.D. & PARRA, E.J. (2000). Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. *Am J Phys Anthropol* **112**, 17–27.
- SWGAT (Scientific Working Group on Materials Analysis) (2005). Forensic human hair examination guidelines. *Forensic Science Communications* **7**(2). Available at http://www.fbi.gov/hq/lab/fsc/backissu/april2005/standards/2005_04_standards02.htm.
- TASI (2004). Colour theory: Understanding and modelling colour. Technical Advisory Service for Images, University of Bristol. Available at <http://www.tasi.ac.uk/advice/creating/pdf/colour.pdf>, accessed December 5, 2007.
- VAUGHN, M., VAN OORSCHOT, R. & BAINBUR-HUDSON, S. (2008). Hair color measurement and variation. *Am J Phys Anthropol* **137**(1), 91–96.
- VAUGHN, M., VAN OORSCHOT, R. & BAINBUR-HUDSON, S. (2009). A comparison of hair colour measurement by digital image analysis with reflective spectrophotometry. *Forensic Sci Int* **183**, 97–101.