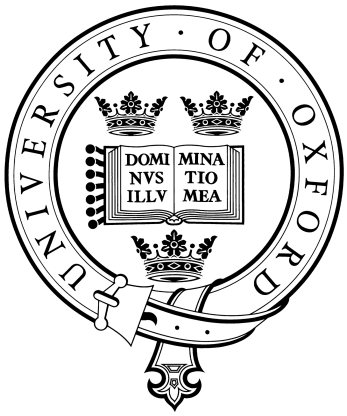
Genetic analysis of normal human phenotypes within the people of the British Isles



Devendra Meena

Green Templeton College

University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2018

Genetic analysis of normal human phenotypes within the people of the British Isles

Devendra Meena
Green Templeton College
D. Phil Thesis, Hilary Term 2018

Abstract

Genome-wide association studies have been of fundamental use in identifying common risk alleles with modest effects underlying multifactorial diseases such as diabetes, cancer, Crohn's disease as well as in polygenic traits, such as pigmentation, involving multiple genes. This thesis, using the GWAS approach, investigates the genetic basis of normal human traits such as human basal pigmentation, sensory perception and food preferences in a well-characterised population of the British Isles, with an emphasis on searching for variants with biologically, as well as statistically, significant effects.

The data used in producing this text were collected as part of the "People of the British Isles" project. Out of 3702 samples, 2,912 were genotyped on the Illumina Human 1.2M-Duo genotyping chip (approximately 1,200,000 SNPs) as part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) studies and the remaining samples were genotyped on Infinium OmniExpress-24 BeadChip array (approximately 750,000 SNPs) at the Wellcome Trust Centre for Human Genetics (WTCHG), Oxford. A total of 547,862 SNPs overlapped both genotyping chips. After genotype quality-control, there were 3,149 samples and 519,188 SNPs remaining for

further analysis. Out of 3,149 samples, only 1,593 samples had phenotype and genotype data available. Fisher's exact test was used to perform association analysis using PLINK (whole-genome data analysis toolset) and the statistical programming language R. In summary, 1,593 and 519,188 SNPs were used in the final genome-wide association analysis.

Among the key features of this thesis are the population genetics data used in analysing different phenotypes. The sample collection criteria for recruiting participants genetically representative of geographic areas in rural Britain is of fundamental importance as it minimises the risk of spurious genotype-phenotype correlations that are due to population stratification. This research, in addition to confirming known genetic associations, presents a number of suggestive novel variants associated with hair and skin pigmentation which influence these complex phenotypes. For normal hair pigmentation, nearly all of the well documented pigmentation genes were replicated, including *MC1R*, *SLC45A2*, *TYR*, *IRF4*, *HERC2*, *KITLG* and *SLC24A4* and results generally agreed with studies performed in the past. Replication of previous associations validates the PoBI dataset.

Next, I demonstrate the genetic basis of bitter taste perception via identification of variants associated with the perception of phenylthiocarbamide (PTC), thiourea (PROP) and sodium benzoate tasting. Variants in the well-known taste receptor gene, *TAS2R38* (rs1726866 and rs10246939), were successfully replicated in the PoBI data. I then focus on the genetic basis of preference for certain food items carefully selected to study the fondness for sweet, sour, bitter and salty tastes, and report novel genetic variants with strong biological evidence. Although none of the SNPs attained

genome-wide significance for food preference data, some SNPs showed suggestive significance and were well-supported by available literature.

The last part of this thesis provides new insights into the field of olfactory perception and argues for the potential link between neurological disorders and olfactory impairment. Most, if not all, novel variants (suggestive genome-wide significant) presented in this research have strong functional biological support behind the preliminary statistical observations and will be helpful in understanding the visible pigmentation differences among humans as well as the genetics underlying food preferences.

In summary, I have taken advantage of a sampling criterion of having four grandparents from the same rural area, minimizing the risk of spurious genotype-phenotype correlations that are due to population stratification. Emphasis on careful collection and selection of the appropriate phenotypic categories has led to results that are both statistically and biologically persuasive, with large genetic effects.

Acknowledgements

First and foremost, I would like to acknowledge my scientific supervisor, Prof. Sir Walter F. Bodmer, for giving me an opportunity to work on the People of the British Isles project and for his guidance and suggestions in producing this research.

I would also like to thank my co-supervisor, Dr. Daniel J. M. Crouch, who has been of great support during the course of my D.Phil. It would not have been possible for me to produce this research without the day-to-day discussions with him. In particular, I must acknowledge his contributions during the final few weeks to get this thesis done within the timelines.

I would also like to thank Dr. Bruce Winney for his support and supervision during the first year of my D.Phil. Many thanks are due to my colleagues Tammy Day and Katarzyna Hutnik for their support during all those years. I wish to thank all the volunteers who took part in the People of the British Isles project, as without their involvement this work would not have been possible.

I would like to express my gratitude to Government of India (GoI) for the financial support during this period. Finally, this work is dedicated to my mother Shrimati Prakashi Devi (who could not attend school due to financial constraints) and my late father Shri Hajari Lal.

Table of Contents

1 Introduction 1

1.1 The Human genome and genetic variation 1

1.1.1 Single Nucleotide Polymorphism 3

1.1.2 Genetic variants and phenotypic diversity 5

1.1.3 Common disease and Common Variant (CD-CV) hypothesis 5

1.1.4 SNPs and Linkage disequilibrium 6

1.1.5 Genome-Wide Association Studies (GWAS) 9

1.2 The People of the British Isles (PoBI) 13

1.3 Human pigmentation 16

1.3.1 Biology of human pigmentation 16

1.3.2 Adaptation of human pigmentation 20

1.3.3 Human pigmentation diversity and GWAS 21

1.4 Population genetics of taste and olfaction 25

1.4.1 The biology of taste and olfaction 26

1.4.2 Olfaction 30

1.5 Genetics of individual food preferences 33

1.6 Summary 34

2 Quality Control 35

2.1 Introduction 35

2.1.1 Sex inconsistencies 37

2.1.2 Sample and marker efficiency 38

2.1.3 Hardy-Weinberg equilibrium 41

2.1.4 Cryptic relatedness 44

2.1.5 Population stratification and batch effects 46

2.2 Discussion 52

2.3 Conclusions 52

3 Population genetics of human pigmentation 53

3.1 Introduction 53

3.2 GWAS of hair pigmentation 55

3.2.1 Introduction 55

3.2.2 Phenotype data collection 56

3.2.3 Statistical analysis 58
3.2.4 Red hair 59
3.2.5 Black versus non-black hair 66
3.2.6 Blond versus non-blond 70
3.2.7 Light versus dark hair 75
3.3 Hair greying 78
3.3.1 *IRF4* and hair greying 78
3.4 Skin pigmentation 85
3.4.1 Melanin-index 88
3.4.2 Erythema-index 89
3.5 Discussion 92
3.6 Conclusions 96
4 Population genetics of taste perception 97
4.1 Introduction 97
4.2 GWAS of bitter taste perception 99
4.2.1 Introduction 99
4.2.2 Phenotype data collection 100
4.2.3 Quality Control 101
4.2.4 GWAS of PTC bitter taste perception 105
4.2.5 GWAS of PROP bitter taste perception 111
4.3 GWAS of sodium benzoate taste perception 116
4.4 Discussion 121
4.5 Conclusions 123
5 Genetics of food preferences 124
5.1 Introduction 124
5.2 Data collection 126
5.3 GWAS of food preference 129
5.4 Discussion 147
5.5 Conclusions 149
6 Genetics of odour perception 150
6.1 Introduction 150
6.2 Phenotype data collection 151
6.2.1 Smell identification test 151

6.3 GWAS of chocolate odour perception.....153
6.4 Smoke odour perception.....158
6.5 Banana odour perception.....163
6.6 Discussion.....166
6.6.1 Protocadherins, autism and sense of smell.....166
6.6.2 Chromosome 15q11-13, olfactory receptors and neurological disorders 167
6.7 Conclusions.....169
7 Discussion.....170
8 References.....178
9 Appendix.....198

List of Figures

Figure 1.1 | Patterns of Linkage Disequilibrium (LD) 8
Figure 1.2 | A composite diagram showing published GWASs. 13
Figure 1.3 | A map of the United Kingdom and sample locations. 15
Figure 1.4 | Human skin colour 16
Figure 1.5 | A hair follicle unit. 17
Figure 1.6 | Diagrammatic presentation of melanin synthesis within a melanocyte 19
Figure 1.7 | A map showing the global distribution of skin colour variation 21
Figure 1.8 | Chemical structures of PTC and PROP 26
Figure 1.9 | Taste buds and receptors 28
Figure 1.10 | Molecular mechanism of taste perception 29
Figure 1.11 | A diagrammatic representation of the vertebrate olfactory system 31
Figure 1.12 | Signal transduction in the olfactory receptor neurons 32
Figure 2.1 | Histogram of X chromosome specific inbreeding coefficient F 39
Figure 2.2 | Genotype and sample missingness 40
Figure 2.3 | Excluding SNPs based on HWE 42
Figure 2.4 | Estimation of IBD from the genotype data. 45
Figure 2.5 | Principal Component Analysis. 47
Figure 2.6 | Principal component analysis 48
Figure 2.7 | Principal component analysis 49
Figure 2.8 | Region-wise distribution of the samples used in the current analysis 50
Figure 3.1 | A part of the questionnaire form used for hair colour data collection. 56
Figure 3.2 | Distribution of hair colour data. 57
Figure 3.3 | Distribution of the hair greying data 58
Figure 3.4 | Manhattan plot for the GWAS of red versus non-red hair 62
Figure 3.5 | Volcano plot for the GWAS of red versus non-red hair 63
Figure 3.6 | Regional association plot of the *MC1R* region 65
Figure 3.7 | Manhattan plot for the GWAS of black versus non-black hair 68
Figure 3.8 | Volcano plot for the GWAS of black versus non-black hair 69
Figure 3.9 | Manhattan plot for the GWAS of blond versus non-blond hair. 72
Figure 3.10 | Volcano plot for the GWAS of blond versus non-blond hair. 73
Figure 3.11 | Regional association plot of the *KITLG* region on chromosome 12 74

Figure 3.12 | Manhattan plot for the GWAS of light versus dark hair. 76
Figure 3.13 | Volcano plot for the GWAS of light versus dark hair. 77
Figure 3.14 | Manhattan plot for the GWAS of hair greying. 80
Figure 3.15 | Volcano plot for the GWAS of hair greying. 81
Figure 3.16 | Q-Q plot for the GWAS of hair greying. 82
Figure 3.17 | Scatter plots of quantitatively measured skin M and E-indices.....86
Figure 3.18 | Histograms of quantitatively measured M and E-indices. 87
Figure 3.19 | Manhattan plots for the GWAS of skin colour. 91
Figure 4.1 | A part of the form used for taste phenotype data collection. 101
Figure 4.2 | Distribution of PTC taste phenotype in the PoBI data. 102
Figure 4.3 | Distribution of PROP taste phenotype in the PoBI data. 103
Figure 4.4 | Distribution of sodium benzoate taste phenotype in the PoBI data. 105
Figure 4.5 | Manhattan plot for the GWAS of PTC taste perception. 107
Figure 4.6 | Volcano plot for the GWAS of PTC taste perception. 108
Figure 4.7 | Regional association plot of chromosome 7 (q34). 110
Figure 4.8 | Manhattan plot for the GWAS of PROP taste perception. 114
Figure 4.9 | Volcano plot for the GWAS of PROP taste perception. 115
Figure 4.10 | Manhattan plot for the GWAS of sodium benzoate taste perception. ..118
Figure 4.11 | Volcano plot for the GWAS of sodium benzoate taste perception. 119
Figure 4.12 | Regional association plot of chromosome 1 (p36.33). 120
Figure 5.1 | Multi-panel bar plots of the food preference data. 128
Figure 5.2 | Volcano plot for the GWAS of butter preference. 131
Figure 5.3 | Manhattan plots for the GWAS of food preferences. 136
Figure 5.4 | Q-Q plots for the GWAS of food preferences. 137
Figure 5.5 | Volcano plot for the GWAS of Marmite preference. 140
Figure 5.6 | Manhattan plots for food preferences. 145
Figure 5.7 | Q-Q plots for the GWAS of food preferences. 146
Figure 6.1 | The Quick Smell Identification Test. 152
Figure 6.2 | Multi-panel bar plots of the odour perception data. 153
Figure 6.3 | Volcano plot for the GWAS of chocolate odour perception. 156
Figure 6.4 | Manhattan plot for the GWAS of chocolate odour perception. 157
Figure 6.5 | Regional association plot showing rs11632280 on chromosome 15. 158
Figure 6.6 | Manhattan plot for the GWAS of smoke odour perception. 161
Figure 6.7 | Volcano plot for the GWAS of smoke odour perception. 162

Figure 6.8 | Manhattan plot for the GWAS of banana odour perception.....165

Abbreviations

GPCRs	G-protein coupled receptors
GWAS	Genome-Wide Association Study
HWE	Hardy–Weinberg equilibrium
IBD	Identity-by-Descent
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
OB	Olfactory Bulb
PoBI	People of the British Isles
PROP	6-n-propylthiouracil
PTC	Phenylthiocarbamide
QC	Quality Control
SNP	Single Nucleotide Polymorphism
TAS2R38	Taste receptor 2 member 38
WTCCC	Wellcome Trust Case Control Consortium

1 Introduction

1.1 The Human genome and genetic variation

The Human Genome Project (HGP), The International HapMap project and the 1000 Genomes Project (Altshuler et al., 2005; Altshuler et al., 2015; Lander et al., 2001) have helped us to advance understanding of the genetic basis of many normal phenotypes and common diseases. The advancement in technology, i.e. genotyping and computing platforms, has made it possible to scan the human genome for hundreds of thousands of genetic markers to detect potential susceptibility loci in a particular disease or phenotype.

Human genetic variation, in a broad sense, can be divided into a) Single nucleotide variants (SNVs) and b) Structural variants (Frazer et al., 2009). Single nucleotide variants (SNVs) are nucleotide changes where only a single base is altered while insertion-deletion (indel, presence of one or more base pairs in some genome while not in others); block substitution (a string of adjacent variants differ in two genomes) or copy number variations (repetition of identical or nearly identical sequences in some chromosome but not in others) are types of structural variants (Frazer et al., 2009). Base substitution and insertion-deletion (indel) of a single base are the fundamental mechanisms underlying the SNP diversity in the human genome. Base substitution can be further divided into two parts: 1) Transition – when a pyrimidine is exchanged for pyrimidine or a purine is exchanged for purine and 2) Transversion - when a pyrimidine/purine is exchanged for a purine/pyrimidine (Jobling, 2014b).

Proofreading errors during DNA replication and mutagenesis either by chemical modification of bases, or physical damage, cause base substitution in the human genome (Jobling, 2014b). SNVs have been extensively used to investigate the correlation between genetic variation and phenotypic differences in human populations.

Unlike SNVs, copy number variants involve large stretches of DNA which are at least 1 kb in size and differ in terms of how many copies of them are situated adjacently on a chromosome compared to a reference genome (Frazer et al., 2009; Redon et al., 2006). In general, CNVs are defined as DNA segments that is at least 1 kb in size, and is present at a variable copy number in comparison with a reference genome (Redon et al., 2006). CNVs can result from insertion-deletion, duplication or triplication (extra copies of a chromosomal region) and translocation (rearrangement of a DNA segment between non-homologous chromosomes) and may account for ~13% of the human genome (Stankiewicz & Lupski, 2010). CNVs can be inherited or a result of genetic mutation and can cause genetic diseases or increase the risk of a complex disorder (Thapar & Cooper, 2013). For example, copy number variation in the human beta-defensin 2 (HBD-2) gene is associated with increased susceptibility to Crohn's disease (Fellermann et al., 2006). Due to the shortage of molecular methods to study CNVs, little is known about their role in the phenotypic variation in the human genome, though some studies have shown genomic distributions of CNVs in the human genome which will help in designing association studies to study the role of CNVs in complex disease biology (Conrad et al., 2010; Zarrei et al., 2015). The 1000 Genomes Project data and advancement in next-generation sequencing methods will help to study CNVs present in the human genome (Altshuler et al., 2015; Pabinger et al., 2014).

1.1.1 Single Nucleotide Polymorphism

A Single Nucleotide Polymorphism (SNP), the simplest form of genetic variation among individuals, is a single nucleotide change (base substitution / insertion-deletion of a single base) in a person's DNA at a given genomic position. Under the classical definition of polymorphism, if a nucleotide, at a given locus, is found altered in more than 1% of the human population it is classified as a SNP. On average, a single nucleotide change occurs every 1000 base-pairs (Brookes, 1999), although human populations vary in terms of the extent of single nucleotide polymorphism observed in the genome. Sequencing of a Yoruba individual showed that one SNP occurs every 1,006 base-pairs which is significantly higher than a genome of European descent which tend to have one SNP per 1,250 base-pairs (Altshuler et al., 2015; Bentley et al., 2008). The 1000 Genomes Project data have shown that there are ~8 million variants which have allele frequency > 5% (Altshuler et al., 2015).

The public catalogue of human genetic variants, the single nucleotide polymorphism database (dbSNP, build 128) contains 12.8 million genetic variants including structural variations (Sachidanandam et al., 2001), though many of these will have very low minor allele frequency (MAF). These markers may be synonymous, meaning their presence in gene-coding regions does not alter the amino acid sequence, or non-synonymous, meaning that they change the resulting amino acid sequence of a protein. The 1000 Genomes Project data have shown that, in a particular genome, there are ~10,000 – 12,000 synonymous and ~10,000 – 11,000 non-synonymous sites (Genomes Project et al., 2010). On the other hand, the majority, ~93%, of GWAS have identified SNPs which are in non-coding regions (Maurano et al., 2012) and it is difficult to predict their functions. Studies in the past have suggested that these non-coding variants could affect the function of

transcription regulatory elements such as promoters and enhancers (Cookson et al., 2009; Maurano et al., 2012). SNPs have been widely used in GWAS to investigate the genetic basis of phenotypic diversity among human populations.

Genotyping is the process of determining a genotype at a particular locus and there are a number of methods available to genotype SNPs. For example, the polymerase chain reaction (PCR) based restriction fragment length polymorphism method can be used to genotype SNPs which create or disrupt restriction enzyme recognition sites (Jobling, 2014b; Parsons & Heflich, 1997) while high throughput DNA microarrays have been widely used in the past decade to perform genome-wide association studies. Affymetrix (Santa Clara, CA) and Illumina (San Diego, CA) platforms are the two major technologies available and have been widely used for GWAS (Bush & Moore, 2012). These platforms offer different versions of SNP chips which contain ~1 million SNPs. A typical SNP genotyping protocol includes: 1) Target fragment (DNA) amplification which can be done using PCR or PCR-free methods; 2) hybridisation and primer extension and 3) allele specific target identification via techniques such as fluorescence resonance energy transfer or mass spectrometry (Chen & Sullivan, 2003; De Angelis, 1999; Jobling, 2014b).

For example, Illumina Human 1.2M-Duo genotyping (used in the current study), SNP chip from Illumina, uses an ordered array of silica beads representing 50-mer oligonucleotides, and assays ~1.2 million SNPs (Distefano & Taverna, 2011; Jobling, 2014b). This bead-based genotyping assay follows three steps in general: 1) Annealing of the amplified DNA to the silica beads which contains 50-mer oligonucleotides; 2) Termination of primer extension by incorporation of complementary modified dideoxynucleotide and 3) Detection of the hybridised product using fluorescent techniques such as mass spectrometry (Jobling, 2014b).

1.1.2 Genetic variants and phenotypic diversity

The fundamental aim of human genetics is to understand the genetic basis of phenotypic diversity within and between human populations. Africans, Europeans and Asians have noticeable phenotypic diversity in terms of pigmentation of hair, eye and skin colour. In addition to these normal human features, scanning the human genome for risk factors underlying different common and complex disorders like cancer, schizophrenia, Alzheimer's and Parkinson's disease, Type 2 diabetes etc. is of central interest in human genetics. The complex nature of normal and disease traits in human populations has a genetic basis (as demonstrated by twin studies), although phenotypes observed in a population may also be influenced by non-genetic factors or by Gene-Environment (GxE) interactions (Brookes, 1999; Frazer et al., 2009).

1.1.3 Common Disease and Common Variant (CD-CV) hypothesis

Pedigree analysis and positional cloning (linkage analysis) have been used in the past to investigate genetic variants in Mendelian genetic disorders (Altmuller et al., 2001; Iles, 2008). The linkage analysis approach was not very successful in identifying common variants which have modest effect size on genetic disorders, usually polygenic in nature e.g. schizophrenia. Therefore, population-based association studies were proposed based on the assumption that a common disease with a genetic component results from the cumulative impact of the multiple common variants, each of which has small to moderate effect size (Lander, 1996; Risch & Merikangas, 1996). In other words, the common disease common variant hypothesis argues that genetic variants with high allele frequency (common, MAF >5%) and low effect size (low penetrance) contribute more towards common diseases than rare alleles with

larger effects. For example, one of the first empirical evidences supporting the CDCV hypothesis was that Apolipoprotein E (APOE) type 4 was found to be associated with the late onset of Alzheimer disease (Corder et al., 1993; Lander, 1996) which was then replicated in a GWAS meta-analysis (Lohmueller et al., 2003).

On the other hand, the CDCV hypothesis has been an issue of debate and it has been argued that rare variants underlie common diseases (Bodmer & Bonilla, 2008). For example, it was suggested that schizophrenia, one of the most complex diseases, is caused by multiple rare alleles (McClellan et al., 2007). On the other hand, for quantitative traits, such as height, which is estimated to be highly heritable (80%), only 5% of the phenotypic variance can be explained by genetic variants (at least 40 loci have been identified) discovered using the GWAS approach (Manolio et al., 2009; Visscher, 2008). Therefore, the CDCV common genetic variants so far fail to explain the genetic heritability of complex diseases (Manolio et al., 2009; Visscher et al., 2012). Despite the criticism of the CDCV hypothesis, genome-wide association studies (GWAS) have been widely used, during the last decade, to dissect the genetic basis of multifactorial diseases like cancer and type 2 diabetes (Chang et al., 2014; Visscher et al., 2012).

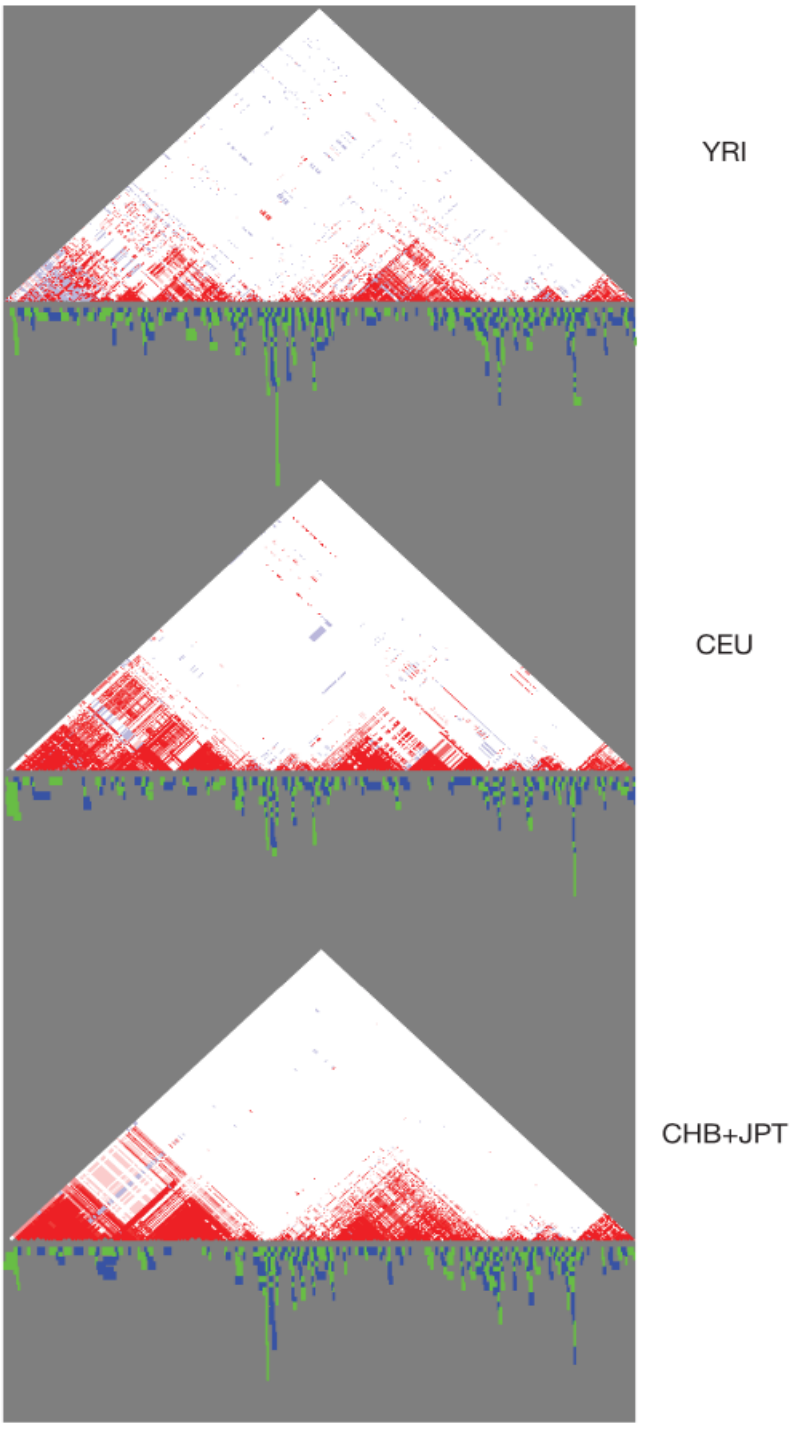
1.1.4 SNPs and linkage disequilibrium

Linkage disequilibrium (LD), a term first used by Lewontin and Kojima (Lewontin & Kojima, 1960), is the association between a pair of alleles, present on neighbouring loci in the genome, because of the recombination rate between the alleles, and is of central interest in human genetics (Tomlinson & Bodmer, 1995). LD, which is usually population-specific, is the association between two genetic markers in a genome and

can be used to study evolution and population history (Slatkin, 2008). The rate at which LD decays with the genomic distance between loci depends on several factors such as selection, random genetic drift, recent population admixture and mutation (Tomlinson & Bodmer, 1995). Thus, human populations differ in terms of degrees and patterns of LD. For example, higher rates of recombination in African population has led to the lower levels of LD in Africans while European and Asian populations have relatively large regions of high LD (Bush & Moore, 2012).

As it is clear from the Fig. 1.1, LD decays with recombination rate, and higher recombination results in discontinuity in LD blocks (Altshuler et al., 2005). "Recombination hot-spot" is a genetic term which describes the genomic regions with elevated rates of recombination. The HapMap project (Phase II) identified 32,996 such recombination hot-spots in the human genome (International HapMap et al., 2007).

Figure 1.1 | Patterns of Linkage Disequilibrium and recombination rate in a 500 kb region of chromosome 2q37.1 in three human populations. The blue and green lines, below each plot, show the adjacent intervals where distinct obligate recombination events must have occurred (Altshuler et al., 2005).



YRI; Yoruba in Ibadan, Nigeria, CEU; Northern Europeans from Utah, CHB; Han Chinese in Beijing, China, JPT; Japanese in Tokyo, Japan.

Population association studies are a form of linkage analysis where LD between two neighbouring alleles can be used to study different multifactorial genetic diseases (Bodmer, 1973; Tomlinson & Bodmer, 1995) and identification of LD patterns formed the main basis for widely used SNP chips in current GWAS studies. The availability of data on the LD patterns in the human genome across over 3.1 million of SNPs from the International HapMap Project made population-based association studies possible (Altshuler et al., 2005; Sachidanandam et al., 2001). The idea of using a *tag* SNP, one that provides information about other unobserved SNPs at nearby sites in the genome due to their alleles being in LD, made it possible to perform association studies by using a subset of SNPs where a few hundreds of thousands of tag SNPs provide high coverage of the genome (Frazer et al., 2009). It is important to note that use of tag SNPs in association studies provides indirect association in almost all cases. In other words, tag SNP is usually not the causal variant and may require further studies to map and validate the causal variant.

1.1.5 Genome-Wide Association Studies (GWAS)

The availability of SNP data, LD patterns in the human genome and SNP genotyping technologies have made it possible to assay hundreds of thousands of markers distributed across the genome, to investigate common genetic variants underlying complex traits. A typical GWAS aims to collect many samples, both 'cases' who have a particular trait of interest, and 'controls', who do not have the trait. Allele frequencies of the genetic variants (SNPs) are then compared between these two groups in order to identify evidence for genetic risk variants underlying the trait of interest. A careful consideration of different factors which could influence the statistical power to capture a causal variant which underlies a trait is important. For

example, definition and selection of phenotype and cases, sample size and a careful selection of tag SNPs is important in design of a GWAS (Klein, 2007; Sham & Purcell, 2014; Spencer et al., 2009). In general, genotyping more individuals on fewer genetic markers will have more power to detect a causal variant compared to genotyping small sample size on more genetic markers (Klein, 2007). In the current study, power calculations were not performed due to time constraints.

In most cases, the SNPs that have lowest p-values and relatively high odds ratios are considered to be significant (Jakobsdottir et al., 2009). "Odds ratio", a term widely-used in epidemiology and genetic association studies, is a measure of effect size defined as the ratio of the odds of carrying an allele for individuals in a group with the disease or trait and equivalent odds in another group without the disease/trait (Clarke et al., 2011; Manolio, 2013). Another measure of a genetic association is the p-value: a stringent significance threshold of 5×10^{-8} is generally considered to correct for multiple corrections in a GWAS study (Fadista et al., 2016).

It can be argued that linkage and candidate gene studies are not the optimal method to understand the polygenic non-Mendelian inheritance in complex diseases such as schizophrenia (Hirschhorn & Daly, 2005) and the availability of a map of human genetic variation made it possible to have hypothesis-free studies in the form of GWAS. Although a study of age-related macular degeneration (AMD) was one of the first genome-wide scans (Klein et al., 2005), shortly afterwards the Welcome Trust Case Control Consortium (WTCCC, 2007) laid the foundation of GWAS by testing ~14,000 cases and 3,000 controls for seven major common diseases - bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 diabetes, and type 2 diabetes (Burton et al., 2007). Since then, and as of March 2017, according to the European Bioinformatics Institute (GWAS catalogue,

http://www.ebi.ac.uk/gwas/, Figure 1.2) 32,320 unique SNP-trait associations have been reported in 2,794 publications (MacArthur et al., 2017).

As discussed earlier, a GWAS aims to report common variants (SNPs with high frequency in the population, >5%) with small effect sizes reflecting the common disease common variant hypothesis. GWAS have, so far, been of great use in understanding of the genetic variants that influence susceptibility to diseases and common complex traits like human pigmentation and these findings have contributed to other fields such as personalised medicine. For instance, genetic factors involved in the metabolism of different drugs identified through GWAS can help in deciding appropriate doses for individual patients (Bush & Moore, 2012; Cooper et al., 2008). Therefore, GWAS have been quite popular in the last decade and contributed to other fields including personalised medicine, forensic science and the activities of consumer genetics companies like 23andMe (Goddard et al., 2016; Visscher et al., 2017).

Despite its popularity, GWAS has scientific limitations and has been criticized for a number of reasons; 1) the majority of common genetic variants identified through GWAS approaches have modest effect size on the trait of interest and only a fraction of variance could be explained by the genetic components. For instance, a meta-analysis identified 71 (of those 32 were identified before meta-analysis and explain 20% of the heritability) loci associated with Crohn's disease and only 23.2 % of the heritability could be explained using these markers (Franke et al., 2010). In this analysis, ~6,000 cases and ~15,000 controls were tested in the discovery phase and 15,694 cases and 14,026 controls were further tested in the replication phase of the meta-analysis. This shows that conducting a GWAS with a large sample size is not enough to explain the heritability of a particular disease, in this case Crohn's disease. This raises a question – what underlies the unexplained heritability? Is it rare genetic

variants or structural variants such a CNVs? Rare variants (MAF <1% in populations) are not well-tagged by common SNPs (as they are less frequent in the population and lack of a catalogue of these variants) used in the commercial GWAS chips, and are hence hard to identify as being significantly associated (McCarthy & Hirschhorn, 2008). Therefore, GWAS is being criticized for focusing on common variants only.

On the other hand, CNVs have been found to be associated with complex disorders such as schizophrenia (Stefansson et al., 2008) and given the progress made to identify them in the human genome, could provide the answer for the unexplained heritability (Conrad et al., 2010; McCarroll, 2008). Almost all GWASs published have reported odds ratios of below 2 and on average 1.1 - 1.4 (Bodmer & Bonilla, 2008) is partly why they only explain a fraction (often <10%) of the estimated heritability for most of the traits studied using GWAS (Manolio, 2013). However, the effect size of these variants must be so small that it is questionable that those remaining undiscovered they will ever be identified, or be of any scientific use;

2) The idea of using *tag* SNPs based on the LD patterns in GWAS creates the question whether the statistically associated variant is itself a causal variant or not;

3) The statistical methods applied to analyse GWAS data, such as Fisher's exact test Pearson's chi-square test (although Fisher's exact test is somewhat conservative, it provides an exact estimate of significance and is robust when sample size is small (Bigdeli et al., 2014)), rely on the assumption of having independent observations and violation (due to factors such as population stratification or relatedness) of these results in type I error or type II errors and 4) testing hundreds of thousands of markers in a GWAS could result in false positive results due to multiple testing unless very stringent penalties are applied. This reduces statistical power, meaning there could be many associated variants that we are missing by using Bonferroni correction.

Figure 1.2 | A diagram showing genome-wide associations ($p \leq 5.0 \times 10^{-8}$) for 17 trait categories which are shown by different colour-coded spots on different chromosomes (http://www.ebi.ac.uk/gwas/) (MacArthur et al., 2017).



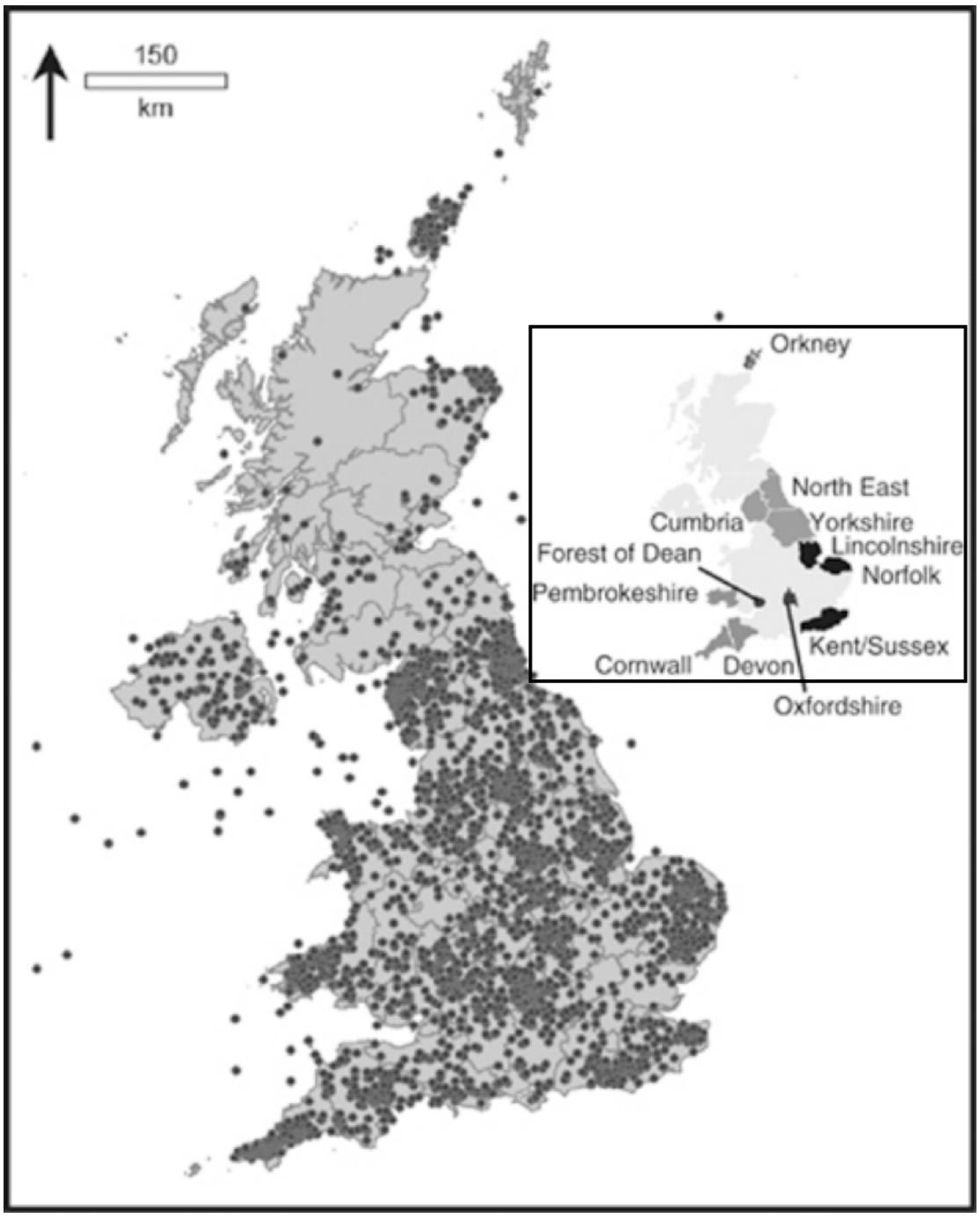
1.2 The People of the British Isles (PoBI)

The People of the British Isles (PoBI) project, funded by the Wellcome Trust, and running since 2004, has collected ~4,500 samples from rural areas of the British Isles with an aim to investigate its fine-scale genetic structure (Leslie et al., 2015; Winney et al., 2012), so as to investigate the migratory history of the British Isles and to provide a well-characterised, carefully collected control population for future studies (Leslie et al., 2015).

Subjects were recruited according to the criteria of having four grandparents born in rural Britain within ~80 km distance from each other (Leslie et al., 2015; Winney et al., 2012). The idea of recruiting volunteers having four grandparents born in the same area of the rural Britain was of particular interest as: 1) Rural areas probably had little migration before 19th century (may be due to factors such as transportation, communication, cost etc.); 2) The grandparents' average age of birth is 1885 (Leslie et al., 2015) so getting all four grandparents from the same rural area means an individual is genetically representative of that area in and before the 19th century and 3) These well-characterised data could be used as a control dataset in medical genetics to study complex disorders, in addition to studying population history.

Of all the subjects recruited as part of the project, most (~3,650 samples) fit the project criteria of having four grandparents from the same rural area (Figure 1.3) (Winney et al., 2012). Although the primary aim of PoBI was to investigate fine-scale genetic differences in the British people, data were also collected, after 2009, to scan for the genetic factors underlying normal human features including pigmentation such as hair; skin and eye colour and to study genes underlying individual taste; food and olfactory preferences.

Figure 1.3 | Distribution of mean geographic position of grandparental birthplaces of the 3,646 volunteers for whom there was information for all four grandparents. Dots (on and outside the map) mark the mean geographic position for individual volunteers. The populations from which samples were taken for genotyping are marked on the inset map (Winney et al., 2012).



1.3 Human pigmentation

Human pigmentation, which is one of the most variable phenotypes among human populations, has been of central interest in cell biology, genetics and anthropology and impacts on our understanding of the evolutionary history of *Homo sapiens*. Attempts to study the genetic basis of hair; skin and eye colour variations in human populations go back to the early twentieth century when researchers attempted to define the inheritance pattern of these traits (Davenport & Davenport, 1907; Gertrude C. Davenport, 1910). Figure 1.4 shows the striking variations in human skin colour.

Figure 1.4 | Variation in human pigmentation apparent through skin colour. An alignment of forearms from lightest to darkest skin colour illustrating a range of skin types (Sturm, 2009).

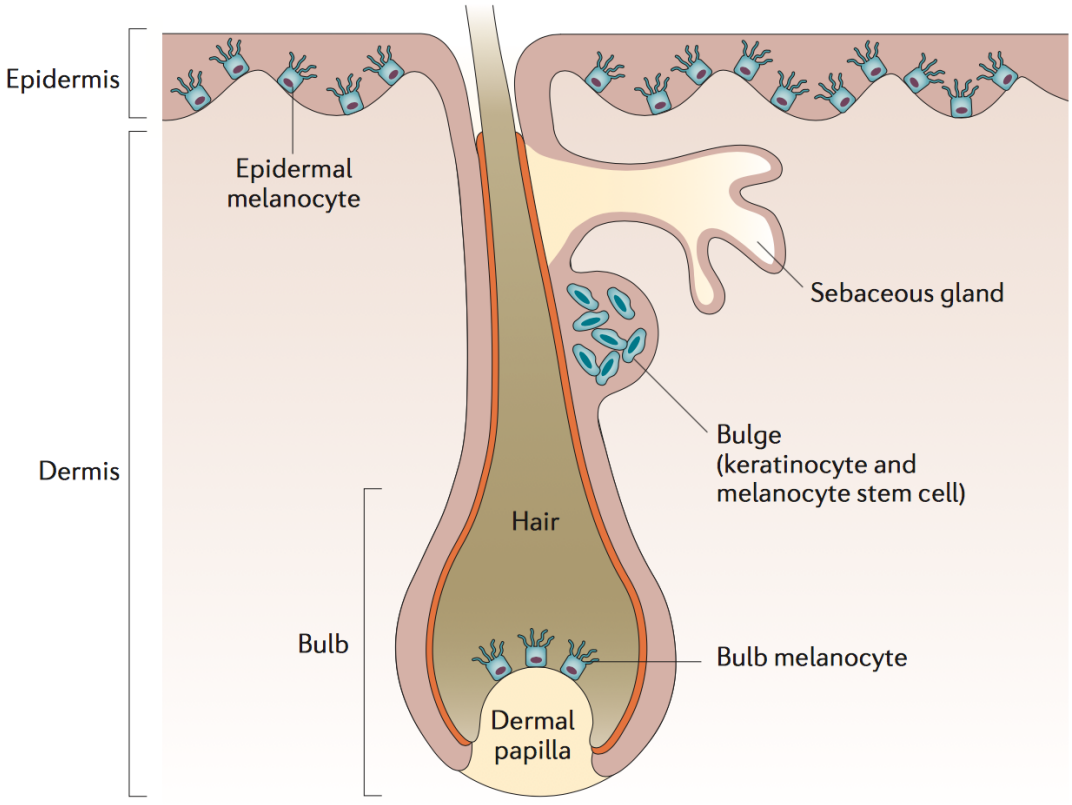


1.3.1 Biology of human pigmentation

Melanin, produced in the melanocytes from tyrosine, is the key molecule in human pigmentation. Melanocytes, present in the basal layer of the epidermis and hair follicle (HF), synthesise melanin, and the amount, type and distribution of this

compound in the melanosomes (pigment granules) determine the visible colour differences among human populations (Sturm, 2009). Here, most importantly, the type of melanin produced in the melanocytes, which could be red/yellow pheomelanin or brown/black eumelanin, is the main factor in visible colour variations in humans (Lin & Fisher, 2007; Sturm et al., 2001). At the cellular level, the origin and development of melanocyte lineage cells is a complex process. At first, regulation of neural crest cells by fibroblast growth factor-2 gives rise to melanocyte lineage cells and these melanocyte precursors then migrate to the dermis and epidermis (Tobin & Paus, 2001). Melanocyte stem cells present in the bulge (Figure 1.5) undergo molecular changes to become melanin producing melanocytes and form the basis of pigmentation differences (Picardo et al., 2015; Sturm, 2009).

Figure 1.5 | The hair follicle unit (Picardo et al., 2015).



While skin colour is maintained by the uniform distribution of melanocytes in the basal layer of epidermis, hair bulb is the only source of pigmentation in a hair follicle unit (Slominski et al., 2005). Although genetic factors that control skin and hair colour have been studied in different populations (Rees & Harding, 2012), what causes hair greying is not fully understood and it is only recently that scientists have claimed to have identified genetic variant that controls this trait (Adhikari et al., 2016). In addition, a study has also suggested that incomplete maintenance of the melanocyte cells in the hair follicle causes hair greying, though the molecular mechanism is not very clear (Nishimura et al., 2005; Steingrimsson et al., 2005).

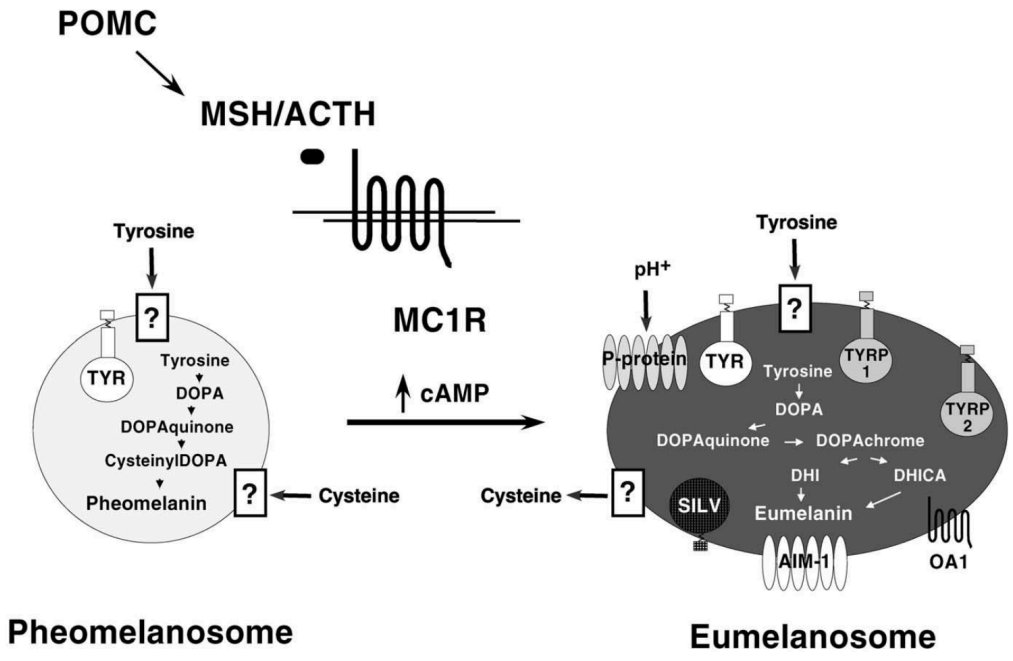
At the molecular level, the melanocortin-1 receptor (MC1R), a member of the G-protein coupled receptor family, plays a key role in mammalian skin and hair pigmentation. MC1R, located on the surface of the melanocytes, interacts with α-melanocyte-stimulating hormone (α-MSH) and adrenocorticotropic hormone (ACTH) to produce eumelanin. As shown in the Figure 1.6, proopiomelanocortin (POMC) cleavage results in the production of α-MSH and ACTH (Sturm et al., 2001). Once activated by α-MSH, MC1R stimulation leads to the increased concentration of the intracellular cyclic AMP (cAMP) which results in synthesis of eumelanin.

The interaction of α-MSH with MC1R results in the activation of downstream signalling molecules such as MITF (microphthalmia-associated transcription factor) in melanocytes (not shown in the figure), which ultimately, results in transcriptional regulation of enzymes, including tyrosinase (TYR) and tyrosinase-related protein 1 (TRP1) - the key regulators in the biochemical synthesis of melanin (Lin & Fisher, 2007; Rees, 2003; Robbins et al., 1993). On the other hand, agouti signal protein

(ASIP) can serve as antagonist and block α-MSH binding to MC1R, therefore dysfunctional MC1R receptor results in low levels of intracellular cAMP which ultimately leads to incorporation of cysteine and retention of sulphur after the synthesis of DOPAquinone producing red/yellow pheomelanin (Wolf Horrell et al., 2016). Variation in the same receptor, MC1R, has been shown to be associated with red hair phenotypes and decreased tanning ability (Valverde et al., 1995).

Although *MC1R* is the major determinant of pigmentation, human pigmentation is a complex trait and different genetic factors at different levels, both cellular and molecular, contribute towards visible colour differences within and between human populations (Sulem et al., 2008).

Figure 1.6 | Diagrammatic presentation of melanin synthesis within a melanocyte. Both, eumelanin and pheomelanin, are derived from a common precursor tyrosine via tyrosinase-dependent pathway (Sturm et al., 2001).



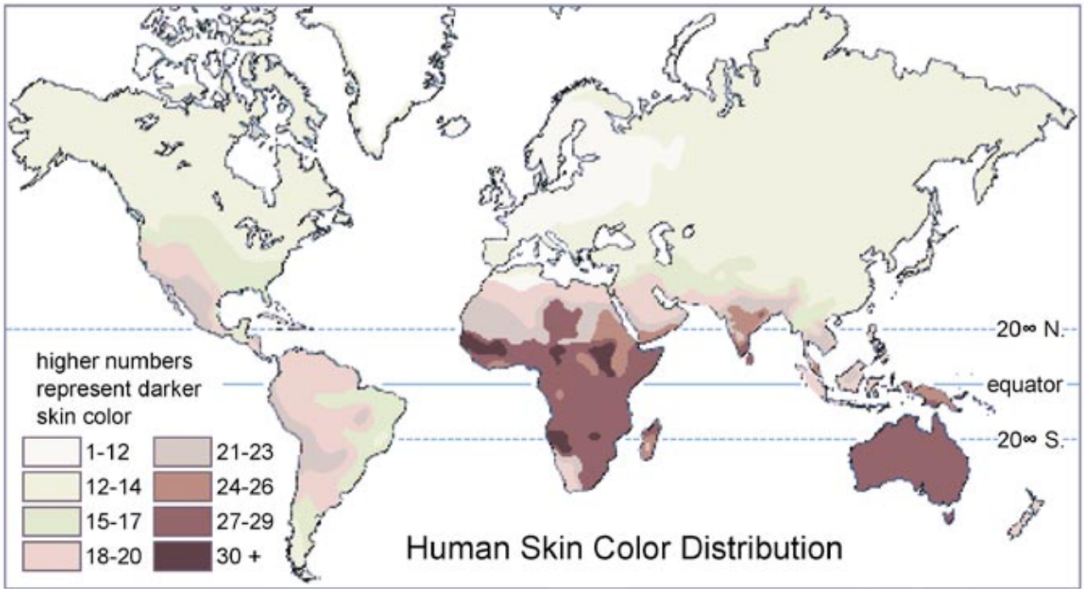
POMC;pro-opiomelanocortin, α-MSH;α-melanocyte-stimulating hormone, ACTH;adrenocorticotropic hormone, cAMP;cyclic adenosine monophosphate, MC1R;melanocortin-1 receptor, TYR;Tyrosinase, TYRP1;Tyrosinase-related protein-1, TYRP2;Tyrosinase-related protein-2, DOPA;3,4-dihydroxyphenylalanine, OA1;ocular albinism type 1, AIM1;absent in melanoma-1.

1.3.2 Adaptation of human pigmentation

Human pigmentation, and skin pigmentation in particular, varies to a great extent between populations and to understand the underlying mechanisms in pigmentation has been an active area of research. Some of the questions are: - what is the basis of this diversity in hair, skin and eye colour among humans? Is it completely genetic or did demographic isolation and adaptations lead us to this great diversity? These remain under discussion and studies done in the past have provided new insights in this field. Two of the proposed theories for the diversity in human skin pigmentation are: 1) Adaptation of eumelanin-rich pigmentation for protection against ultra-violet (UV) radiation as UV radiation can induce DNA damage and 2) Adaptation of light skin pigmentation for increased absorption of UV radiation as UV radiation is important to synthesise vitamin D3 (Jablonski & Chaplin, 2000, 2010). In addition to the above-mentioned theories, an ancient DNA study has shown that light skin pigmentation in Europeans was already present in the Bronze Age (Allentoft et al., 2015).

As shown in the Figure 1.7, human populations close to the equator tend to have dark skin colour for protections against UV radiation while populations at higher latitude are tend to have light skin colour which is important for vitamin D photosynthesis (Jablonski & Chaplin, 2010).

Figure 1.7 | A map showing the global distribution of skin colour variation (Barsh, 2003).



1.3.3 Human pigmentation diversity and GWAS

Human pigmentation is a polygenic trait. The classical example of abnormal pigmentation comes from the studies of oculocutaneous albinism (OCA) which results from the altered function of melanin synthesising enzyme, Tyrosinase (King et al., 1991). All forms of oculocutaneous albinism (OCA) show a defect in pigmentation production in hair; skin and eyes and mutations (in tyrosinase, *TRYP1*, *OCA2*/P protein and *MAPT*) affecting this phenotype were known even before the GWAS era (Jackson, 1997; Newton et al., 2001).

Although studies, including mutations affecting mouse coat-color identified at least five loci (Jackson, 1997), the majority of them remained unknown until the use of

association studies. New approaches such as genome-wide association studies (GWAS) have significantly contributed towards what we know about these traits (Table 1.1 for a list of GWASs of pigmentation phenotypes). For example, GWASs in Europeans have reported novel associations with pigmentation including *SLC24A4* and *KITLG* (Han et al., 2008; Sulem et al., 2007), but these do not explain the full range of phenotypic variation, which suggests that there could be other genetic factors. Therefore, analysing different populations using novel sampling strategies, such as in the PoBI project, may provide new insights in this field.

Table 1.1 | A list of pigmentation genome-wide association studies.

Phenotype	**Population**	**Sample size (discovery/replication)**	**Gene(s) identified**	**Reference**
Skin colour	South Asian	737/231	*TYR*, *SLC45A2*, *SLC24A5*	(Stokowski et al., 2007)
Hair, skin & eye colour, freckles	Europeans	2,986(Icelandic)/2,718(Icelandic)& 1,214(Dutch)	*SLC24A4*, *KITLG*, *TYR, OCA2*, *MC1R*	(Sulem et al., 2007)
Hair & skin colour, freckles	Europeans	5,130(Icelandic)/2,116(Icelandic)& 1,214(Dutch)	*TPCN2*, *ASIP*	(Sulem et al., 2008)
Hair & skin colour	Europeans	2,287/7,028	*SLC24A4*, *IRF4*, *SLC45A2*, *HERC2*	(Han et al., 2008)
Eye colour	Europeans	1,406	*HERC2*/*OCA2*	(Kayser et al., 2008)
Hair, skin & eye colour	Europeans	176/294	*HERC2*/*OCA2*	(Candille et al., 2012)
Skin & eye colour	African-European	699	*SLC24A5*, *SLC45A2*, *TYR*, *HERC2/OCA2*	(Beleza et al., 2013)

Hair, skin & eye colour	European-Americans	10,183/4,504	*EDNRB*, *VASH2*	(Zhang et al., 2013)
Skin colour	Europeans	17,262	*IRF4*, *SLC45A2*, *MC1R*, *ASIP*, *HERC2/OCA2*	(Liu et al., 2015)
Hair colour	Europeans	7,091/3,340	*MC1R*, *HERC2*, *TPCN2*, *SLC24A4*, *IRF4*, *KITLG*	(Lin et al., 2015)
Hair colour	Latin Americans	6,000	*SLC45A2*, *IRF4*, *TYR*, *HERC2/OCA2*, *SLC24A5*	(Adhikari et al., 2016)
Skin colour	Africans	1,570	*SLC24A5*, *MFSD12*, *DDB1*, *TMEM138*, *HERC2/OCA2*	(Crawford et al., 2017)
Skin colour	African-Americans	285/373	*BEND7/PRPF18*, *SLC24A5*, *SLC45A2*	(Hernandez-Pacheco et al., 2017)

TYR; tyrosinase, *SLC45A2*;solute carrier family 45 member 2, *SLC24A5*;solute carrier family 24 member 5, *SLC24A4*; Solute carrier family 24 member 4, *KITLG*; Tyrosine-protein kinase kit-ligand, *MC1R*;melanocortin-1 receptor, *OCA2*;Oculocutaneous albinism type 2, *HERC2*;HECT and RLD domain containing E3 ubiquitin protein ligase 2, *TPCN2*; Two pore segment channel 2, *ASIP*; Agouti-signalling protein, *IRF4*; Interferon regulatory factor 4, *EDNRB*; Endothelin receptor type B, *VASH2*; Vasohibin 2, *MFSD12*; Major facilitator superfamily domain containing 12, *DDB1*; DNA damage-binding protein 1, *TMEM138*; Transmembrane Protein 138, *BEND7*; BEN domain-containing protein 7, *PRPF18*; pre-mRNA processing factor 18.

1.4 **Population genetics of taste and olfaction**

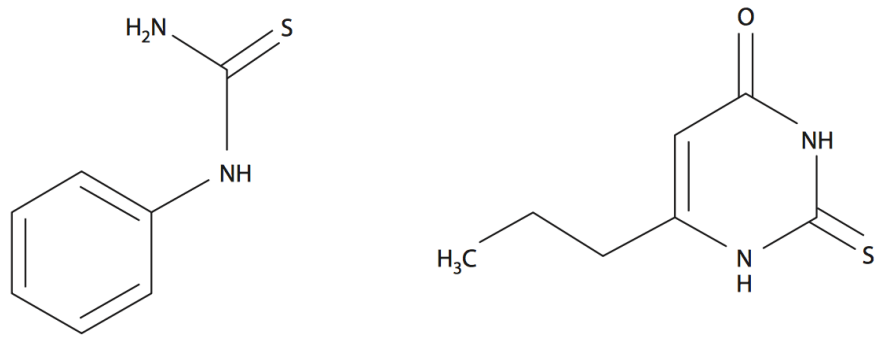
An accidental laboratory incident occurred in 1932 when A. L. Fox was placing phenylthiocarbamide (PTC, Figure 1.8, PROP – a chemically related compound to PTC) into a bottle and the dust (PTC powder) 'flew around in the air', producing some interesting observations. C. R. Noller, tasted the 'dust' as bitter while Fox did not. Fox then proposed that humans vary in the ability to taste phenylthiocarbamide (PTC, an organosulphur thiourea that tastes bitter to most individuals), and thus laid the foundation of taste and olfactory research (Fox, 1932). Soon after Fox's discovery, R. A. Fisher (with J Huxley and E. B. Ford) performed experiments on chimpanzees to test if the variation also occurs among them and he found that chimpanzees also vary in their ability to taste PTC (Fisher, 1939). Since then, understanding the molecular mechanisms of sensory perception has been an active area of research. One of the first reports of the molecular basis of PTC bitter taste perception comes from a linkage analysis performed in the late 90s (Reed et al., 1999) in which it was shown that the ability to taste PTC could be linked to a locus near *KELL* on chromosome 7 (q33).

In addition to taste perception, olfactory perception plays an important role in day-to-day life, and yet the sense of smell is poorly investigated. Since the discovery of olfactory receptors (Buck & Axel, 1991), not many studies have been conducted to identify additional genetic variants underlying sense of smell with some recent exceptions (McRae et al., 2013).

1.4.1 The biology of taste and olfaction

An individual's taste perception ability, at the cellular level, depends on the taste-receptor cells (TRC) distributed in the taste buds on the papillae of the tongue and palate epithelium (Figure 1.9) (Chandrashekar et al., 2006). These TRCs play a key role in the way humans distinguish between the five major components of perceivable flavours: bitter, salty, sour, sweet and umami.

Figure 1.8 | Chemical structures of phenylthiocarbamide and 6-n-propylthiouracil (Keller et al., 2002)



Phenylthiocarbamide **6-*n*-Propylthiouracil**

At the molecular level, members of the G-protein coupled receptors (GPCRs) known as T1Rs and T2Rs mediate taste perception. Taste receptor type 1 (T1R) mediates sweet taste perception (Nelson et al., 2001) while taste receptor type 2 (T2R) mediates bitter taste perception (Chandrashekar et al., 2000). Some studies have also suggested

that a combination of taste receptors mediate amino acid taste (umami) perception (Nelson et al., 2002).

In most cases, upon binding of taste molecules to GPCR on the surface of a taste receptor cell, freed Gα and Gβγ interacts with PLCβ2 (Phospholipase C beta 2) to increase intracellular levels of Ca^{2+} via IP_3R3 (IP3 receptor subtype 3) ion channels (Chaudhari & Roper, 2010) which results in the depolarisation of the cells through the cation channel TrpM5 (Figure 1.10). As an ultimate result of the intracellular levels of Ca^{2+}, release of taste bud transmitter ATP through hemi-channel pores, such as pannexin, in the extracellular space surrounding the taste receptor cells results in taste perception (Chaudhari & Roper, 2010; Huang & Roper, 2010). On the other hand, sour taste stimuli (acids) exert membrane depolarisation directly via H^+ released from organic acids, and the salty taste of Na^+ results from direct depolarisation of taste receptor cells via Na^+ ions which are a feature of salty foods (Figure 1.10) (Chandrashekar et al., 2010; Chaudhari & Roper, 2010).

In addition to T1Rs and T2Rs, there are receptors which mediate sensory perception of different molecules. For example, a recent study has suggested that ability to taste glucose oligomers is independent of T1R receptors (Lapis et al., 2016). Therefore, it is possible that there may be other receptors present in taste receptor cells, which may underlie different aspects of taste perception, but remain undiscovered and may help to explain sour, salty or umami taste perception which are at present poorly understood.

Figure 1.9 | Diagrammatic representation of the taste buds and taste receptor cells on tongue papillae[a]. The bottom part shows that responsiveness to different taste stimulant is not region-specific (Chandrashekar et al., 2006).

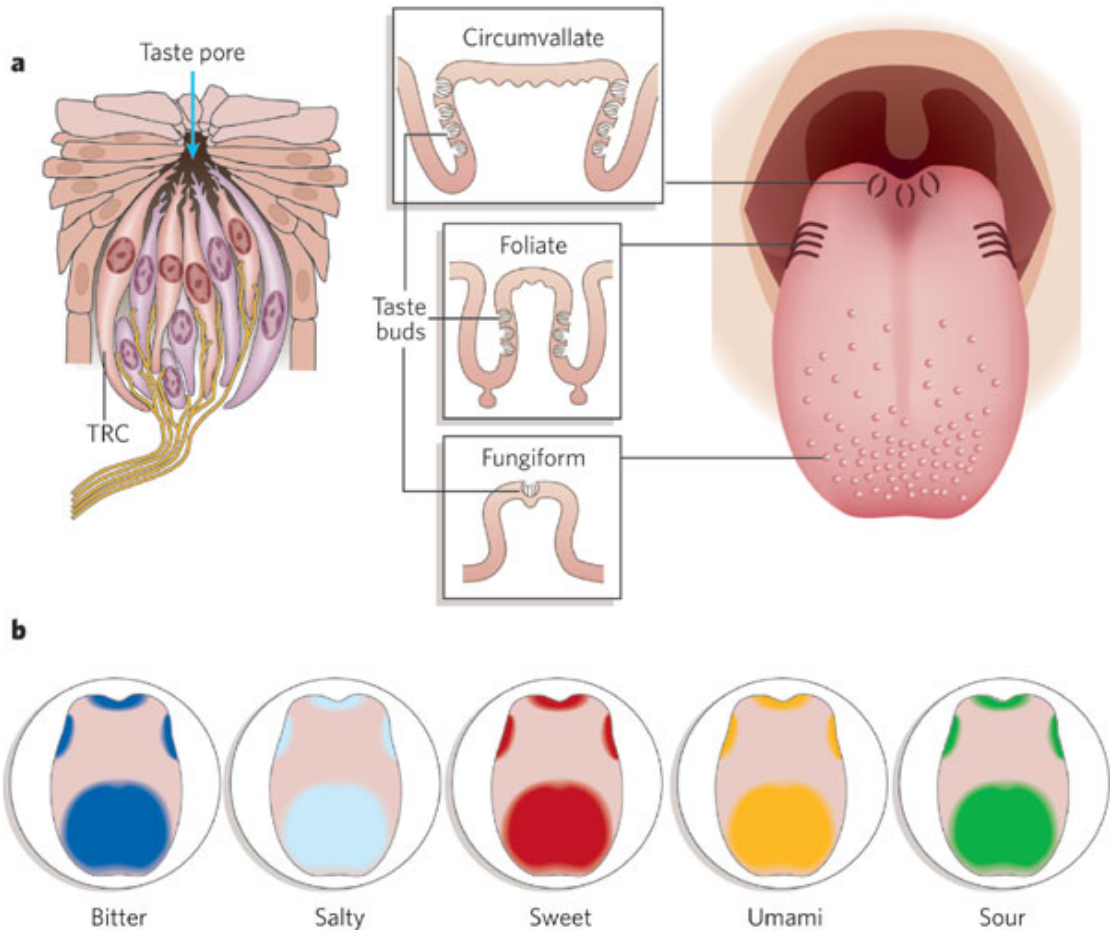
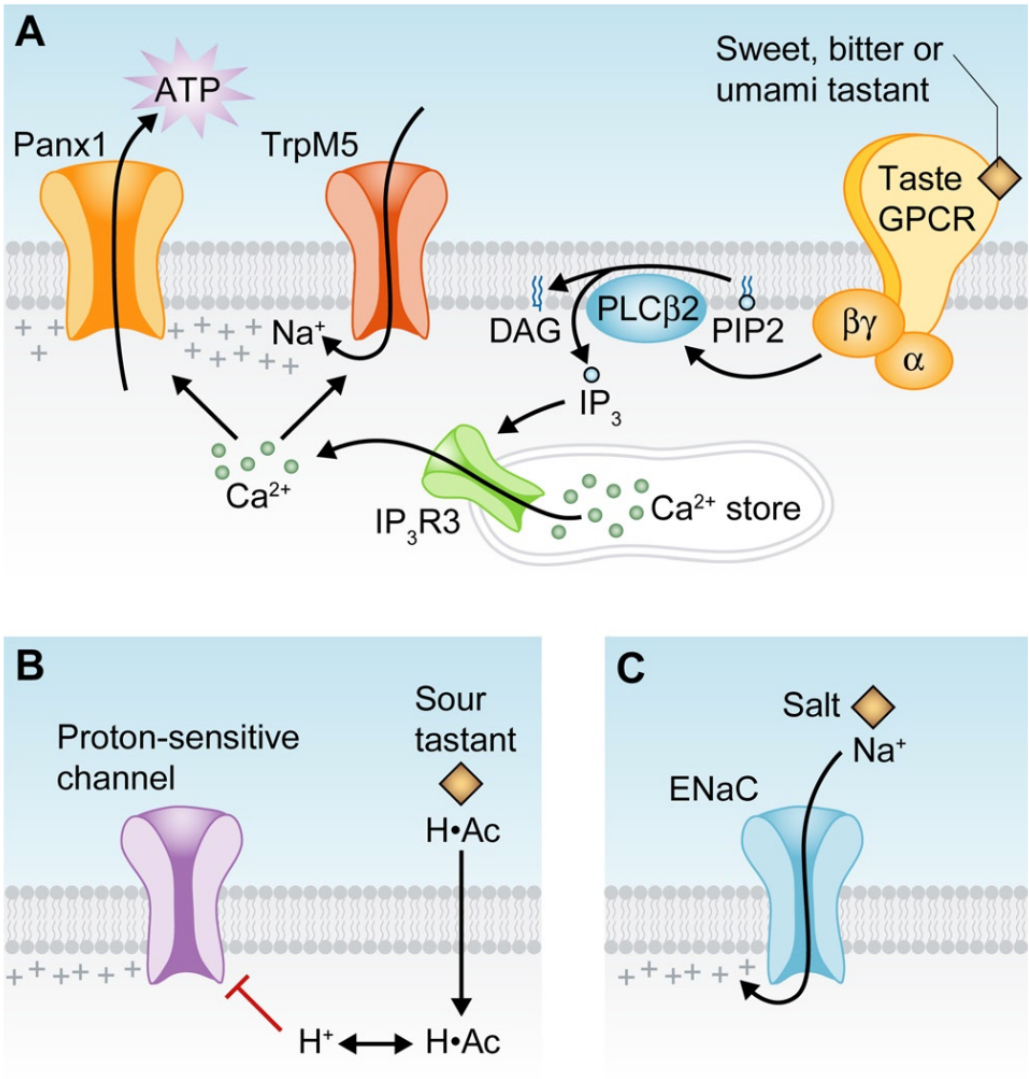


Figure 1.10 | Schematic representation of molecular mechanisms for sweet, bitter, umami, sour and salty taste transduction (Chaudhari & Roper, 2010) .



GPCR;G-protein-coupled receptor, α & βγ;types of G-protein, PIP2;Phosphatidylinositol 4,5-bisphosphate, PLCβ2;Phospholipase C beta 2, DAG;diacylglycerol, IP3;Inositol triphosphate, IP_3R3;IP3 receptor subtype 3, TrpM5;Transient receptor potential cation channel subfamily M member 5, Panx1;Pannexin 1, ATP; Adenosine triphosphate, H.Ac;Organic acids, ENaC;Epithelial sodium channel.

1.4.2 Olfaction

Olfactory sensory neurons, present in the olfactory epithelium (OE) in the nasal cavity, play a fundamental role in olfaction (sense of smell) in humans (Menini et al., 2004). As shown in Figure 1.11, olfactory sensory neurons mediate downstream signalling and pass the sensory input to the glomeruli in the olfactory bulb (OB), which in turn, transmits this information to the olfactory cortex resulting in odour perception (Figure 1.11 and Figure 1.12)(Menini et al., 2004). The olfactory receptors (OR) expressed by olfactory sensory neurons are members of the G-protein coupled receptor (GPCR) family, which mediate an individual's sense of smell (Buck & Axel, 1991). Again, as in taste perception, Ca^{2+} plays a key role in downstream signalling of an odorant stimulus.

Upon binding of odorant to olfactory receptor, the increased intracellular concentration of Ca^{2+} and Na^{2+} causes depolarization of the neurons, which ultimately, results in release of neurotransmitters at the synaptic terminal in the olfactory bulb (Brunet et al., 1996; DeMaria & Ngai, 2010). Studies, including GWAS, have reported new insights suggesting that olfactory perception influences food preferences (Eriksson et al., 2012) and that olfactory-impairment is a potential indicator of the neurodegenerative diseases (Dong et al., 2015). Therefore, chemosensory perception in humans may have molecular links with different neurological disorders, and discovery of new variants could shed light on the direction of future medical research.

Figure 1.11 | A diagrammatic representation of the vertebrate olfactory system (Menini et al., 2004).

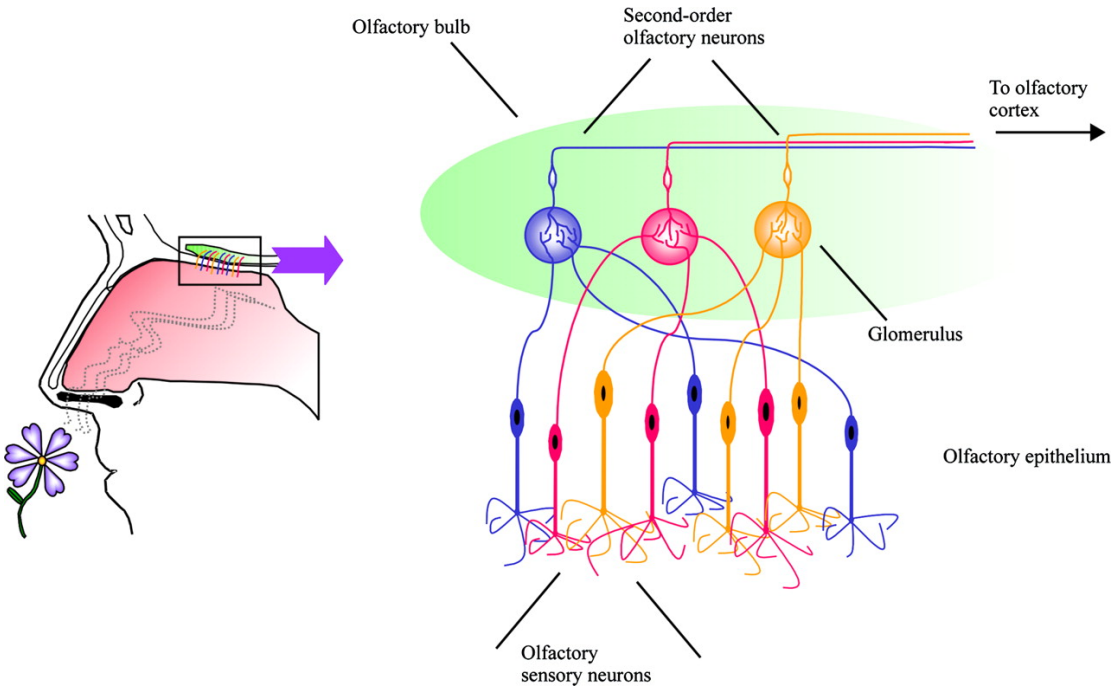
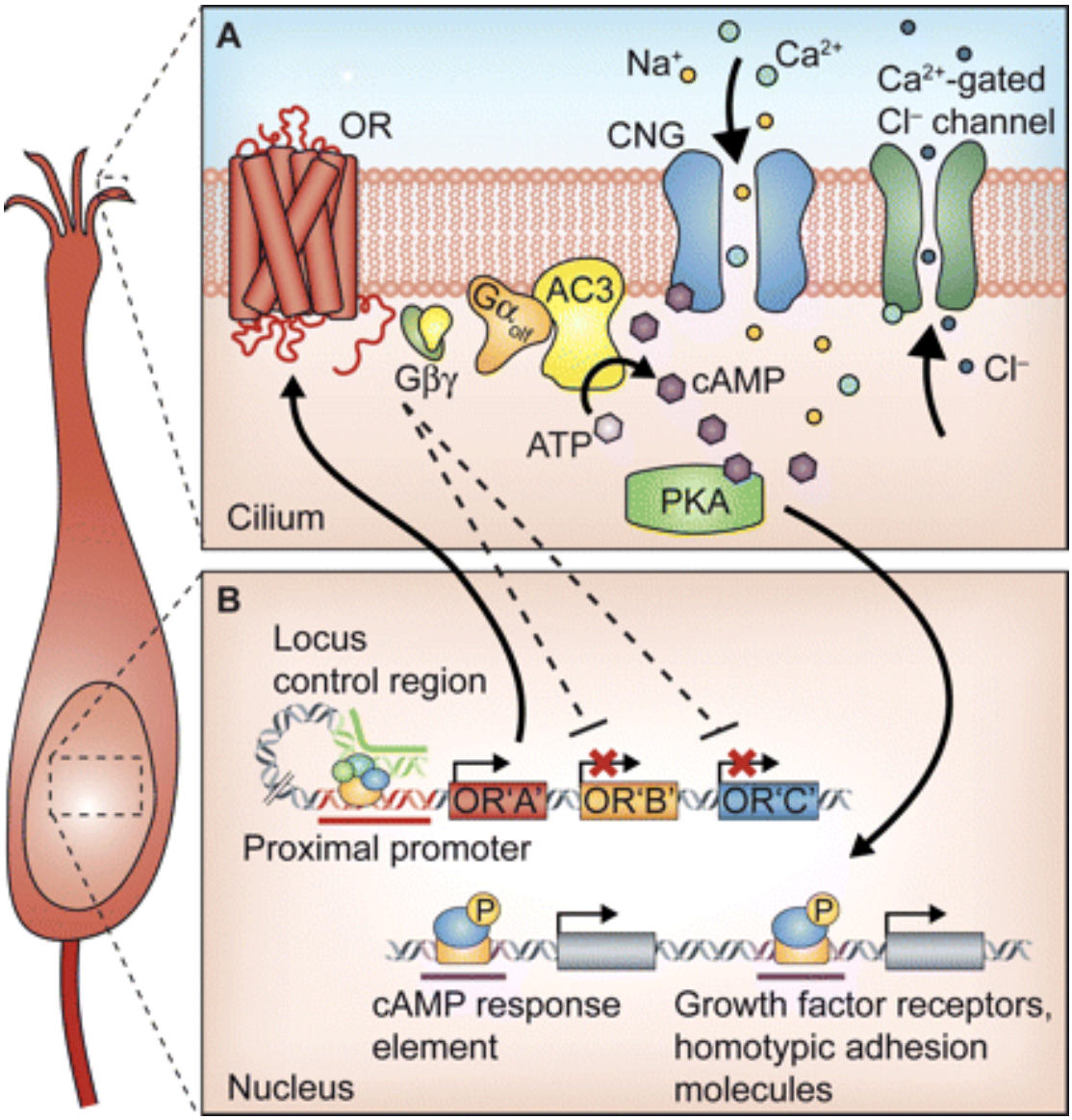


Figure 1.12 | Diagrammatic representation of signal transduction in olfactory receptor neurons (DeMaria & Ngai, 2010).



OR;Olfactory receptor, CNG;Cyclic nucleotide-gated, $G\alpha_{olf}$ & Gβγ;types of G-protein, cAMP;Cyclic adenosine monophosphate, AC3;Type III adenylyl cyclase, PKA;Protein kinase A, ATP; Adenosine triphosphate

1.5 Genetics of individual food preferences

Food preferences vary to a great extent, within and between human populations, and environmental and socio-cultural factors influence food selection (Smith et al., 2016; Ventura & Worobey, 2013). For instance, available resources, family-environment, economic factors or personal experiences could influence an individual's preference for certain foods. On the other hand, the identification of bitter and sweet taste receptors and the great diversity in human olfactory receptors raises the question: are there any genetic factors that mediate an individual's food preference and do chemosensory variations have any influence on food selection? Studies in monozygotic (MZ) and dizygotic (DZ) twins have shown that food preferences have a moderate genetic basis, for example, h^2 of 0.36 for fruit and vegetable preference (Hasselbalch et al., 2008; Pallister et al., 2015). Additionally, a GWAS analysis has shown that white wine liking is influenced by genetic factors (Pirastu et al., 2015). Another study has shown that variation in taste receptors influence the risk of alcohol dependence (Hinrichs et al., 2006) which surely affects individuals' likelihoods of selecting alcoholic drinks.

It is important to note that dietary intake may influence an individual's susceptibility to various diseases, for example cardiovascular disease and obesity (De Caterina et al., 2006). Therefore, there is growing interest in understanding how genetic factors contribute towards an individual's food preferences, and what role these genetic factors play in the development of chronic conditions.

1.6 **Summary**

In summary, the use of genome-wide association studies for identification of genetic variants associated with pigmentation, food preferences, taste and olfactory perception will provide insights into the medical, anthropological and forensic sciences. The main goal of this thesis is to use the GWAS approach to search for novel genetic associations in a well-characterised population.

2 Quality Control

2.1 Introduction

A typical genome-wide association study (GWAS) involves: 1) Study design and definition of cases and controls; 2) data collection (both phenotype and genotype) and quality control; 3) statistical analysis and 4) validation and replication of observed associations in a meta-analysis. Among these, stringent quality control (QC) parameters are important to avoid potential misinterpretation of data (Burton et al., 2007). For instance, observed associations in a GWAS dataset, not accounting for population stratification and cryptic relatedness, could be due to substructure present in the analysed population, rather than genuine causal relationships between alleles and phenotypes. Biological factors such as relatedness and inbreeding, and technical factors such as batch effects and SNP genotyping problems may result in spurious associations. For example, a case-control association study that includes individuals from different ethnicities would produce spurious result due to allele frequency differences among the analysed samples, if the frequency of one ethnicity was higher in cases than it was in controls. A standard quality control (QC) procedure in a GWAS can be divided into several parts. Sample-based QC involves checking each individual for sex discrepancies (against self-reported sex), cryptic relatedness, sample duplicates, population outliers and samples with low call rate, which if not accounted for, may produce spurious associations with the trait of interest. On the other hand, SNP-based QC consists of eliminating SNPs with low call rate, deviation from Hardy-Weinberg Equilibrium (HWE) and minor allele frequency (MAF). It is crucial to exclude samples and SNPs which violate standard QC criteria to avoid type I or type II errors (McCarthy et al., 2008; Weale, 2010). For example, population

structure increases type I error due to causing correlation between alleles and phenotype data that do not represent genetic causation, and sample mismatches (i.e. the genotypes are paired together with the wrong individual IDs and hence the wrong phenotype measurements) will decrease statistical power causing type II error.

Before applying quality control filters, there were 3,702 samples, collected as part of the People of the British Isles (PoBI) project and DNA extraction and genotyping was performed by others as mentioned in the initial publication of this study (Winney et al., 2012). I was involved in the sample collection, and performed the statistical analyses (phenotype data analysis, quality-control and genetic association analysis). Out of 3702 samples, 2,912 were genotyped on the Illumina Human 1.2M-Duo genotyping chip (approximately 1,200,000 SNPs) as part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) studies (Genetic Analysis of Psoriasis et al., 2010) and the remaining samples were genotyped on Infinium OmniExpress-24 BeadChip array (approximately 750,000 SNPs) at the Wellcome Trust Centre for Human Genetics (WTCHG), Oxford. The second chip was used due to financial considerations.

A total of 547,862 SNPs overlapped after combining the data from both genotyping platforms. The statistical analysis for QC and association analysis was performed using PLINK versions 1.07 (Purcell et al., 2007) and 1.90 (Chang et al., 2015) and the R statistical programming language (http://www.R-project.org/). Plots (such as volcano and Q-Q plots) were created using R versions 3.4.2 or earlier (http://www.R-project.org/). The Manhattan and regional association plots were produced using the R package qqman (Turner, 2014) and the web tool LocusZoom (Pruim et al., 2010), respectively. A consistent threshold of 5×10^{-8} and 1×10^{-5} were considered to be

genome-wide significant and suggestive genome-wide significant, respectively. In case the genetic association is within a previously associated SNP, a p-value of 0.05 can be considered significant as it constitutes an independent replication. It is important to note that the final analysis was performed only on autosomal markers as more complex statistical tests are required to analyse the X chromosome (Burton et al., 2007; Clayton, 2009; Wise et al., 2013).

2.1.1 Sex inconsistencies

X and Y chromosome data can be used to check and identify potential sample handling errors, thus helping to confirm that phenotype and genetic data match with each other. This is achieved by using the X chromosome heterozygosity rate to predict sex from genetic data (the "--check-sex" option in PLINK), as hemizygous male genotypes appear as homozygotes on a GWAS array. Based on the X chromosome heterozygosity rate, PLINK assigns sex to a particular sample and checks this with the sex provided in the pedigree file which contains sample IDs and other information such as family status. If discrepancies occur, then PLINK assigns the potential mismatch as a 'problem' (Purcell et al., 2007). PLINK uses the inbreeding coefficient *F,* which describes the extent of excess heterozygosity relative to Hardy-Weinberg Equilibrium, to determine male and female status where *F* should be close to zero for genetic females and close to 1 for genetic males (Figure 2.1). In total, 173 samples were excluded (Table 2.1) based on sex discrepancies where self-reported sex did not match with genetically predicted sex or where the X-chromosome inbreeding coefficient *F* values were between 0.28 to 0.8, implying uncertain sex based on the SNP data. A histogram showing the distribution of the X-chromosome inbreeding coefficient *F* is shown in Figure 2.1 (before and after

excluding the sex discrepancies). There were 3,529 samples left after removing sex discrepancies (Table 2.1).

2.1.2 Sample and marker efficiency

Both samples and SNPs with low rates of successful genotyping could indicate poor DNA quality or technical problems in calling genotypes, and should therefore be excluded from downstream analysis as either of these problems can lead to spurious associations. Missingess in GWAS data can be categorised as either individual missingess or SNP missingness. These are controlled by excluding SNPs and individuals which do not meet certain missingness criteria. Figure 2.2 shows the sample (individual) genotyping call rate and SNP coverage (defined as the proportion of assays which failed for a given SNP) before and after excluding samples and SNPs. Although setting up a threshold for excluding samples and SNPs based on call rate may depend on the study in question, excluding 1-2% genotype missingness rate is typically recommended (Burton et al., 2007; Turner et al., 2011). Samples and SNPs with < 99% call rate were excluded from further analysis (Figure 2.2). After removal, there were 3,432 samples and 537,501 SNPs remaining.

Figure 2.1 | Histogram of the X-chromosome specific inbreeding coefficient *F* before (a) and after (b) removing sex discrepancies. Genetic females have *F* close to 0 and genetic males have *F* close to 1. Intermediate *F* individuals (0.2<F<0.8) are thus excluded from the data. The y-axis has been cropped to better show the counts of *F* values between 0.4 to 1.

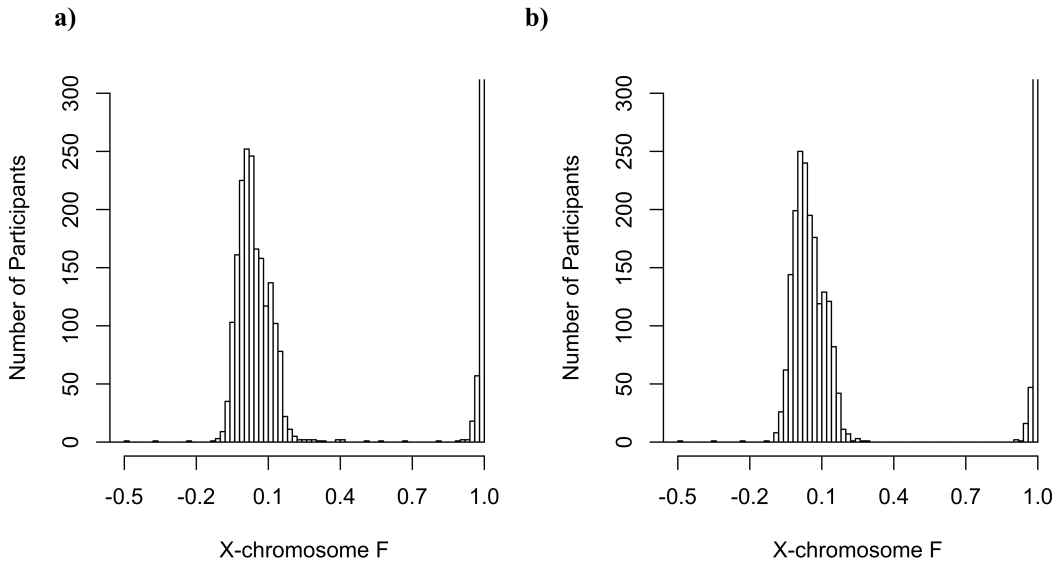
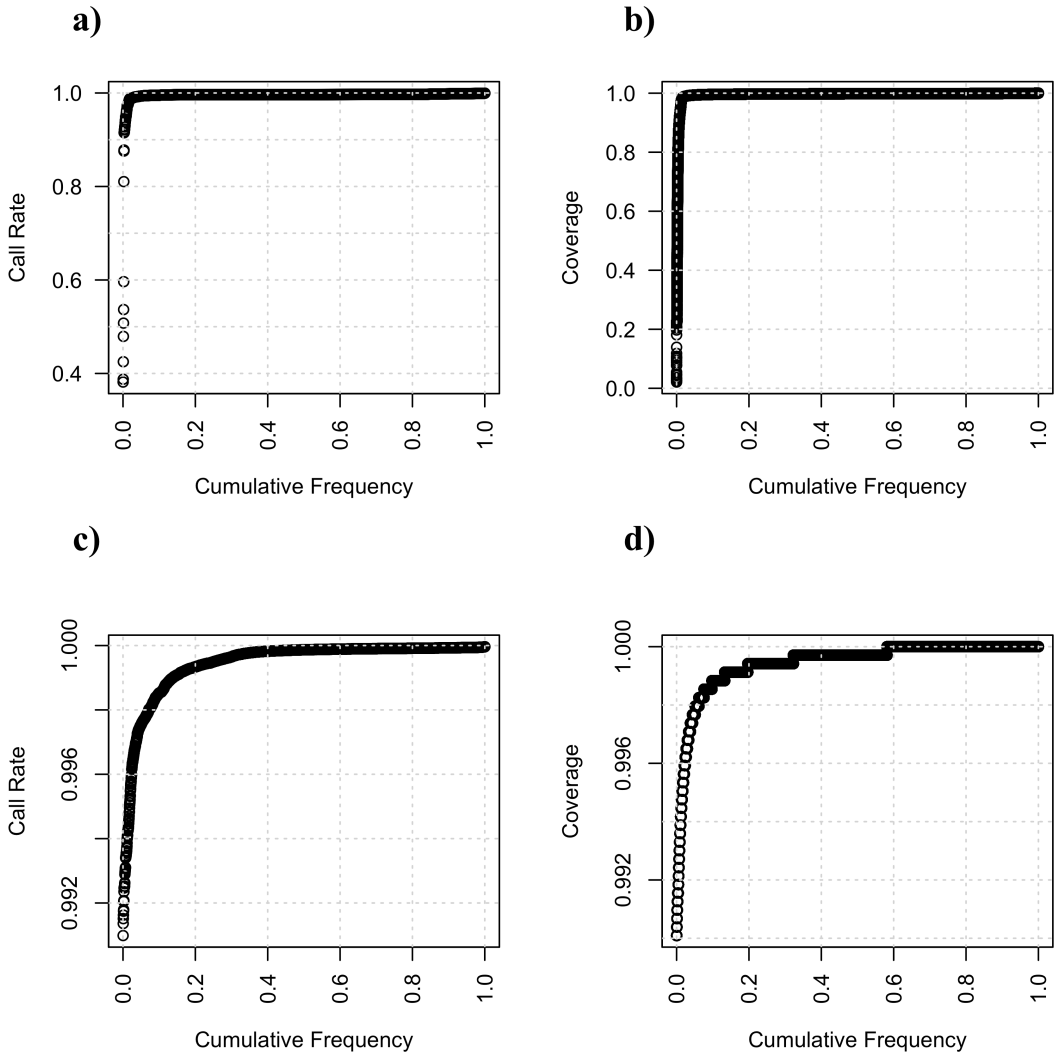


Figure 2.2 | Plots of genotype missingness rate - call rate for samples and coverage for SNPs, against cumulative frequency before (a & b) and after (c & d) implementing the QC threshold. Samples with <99% call rate and SNPs with <99% coverage were excluded from the data.



2.1.3 Hardy-Weinberg equilibrium

Departures from Hardy-Weinberg equilibrium (HWE) could be an indication of population structure (though this is unlikely in the current PoBI dataset due to the sampling strategy), or genotyping errors; therefore, SNPs deviating strongly from HWE should be removed from further analysis (Anderson et al., 2010; Cox & Kraft, 2006). They can also be the result of powerful natural selection (Anderson et al., 2010). It is also important to note that departure of HWE could also represent a true genetic association with the trait under study (Wittke-Thompson et al., 2005). Very extreme departures from HWE are mostly due to genotyping errors (Weale, 2010). Figure 2.3 shows the p-values calculated for deviation from HWE using an exact test (Wigginton et al., 2005). Although cases and controls are often tested separately for this purpose, as they were not collected or genotyped separately, HWE was tested on the whole dataset. A total of 5,413 SNPs (~1% of total SNPs) were removed based on HWE (Figure 2.3 and Table 2.1). Although ~1% of total SNPs seems like quite a large proportion to exclude, a stringent QC is the key to avoid any false-positive results in any GWAS.

Figure 2.3 | Quantile-Quantile (Q-Q) plots of p-values ($-\log_{10}$ transformed) for departure from HWE calculated using an exact test (before and after excluding 5,413 SNPs deviating from HWE). The observed p-values for all SNPs were plotted against the expected p-values. A clear deviation (before QC) of the observed p-values from the expected represents SNPs departing from HWE.

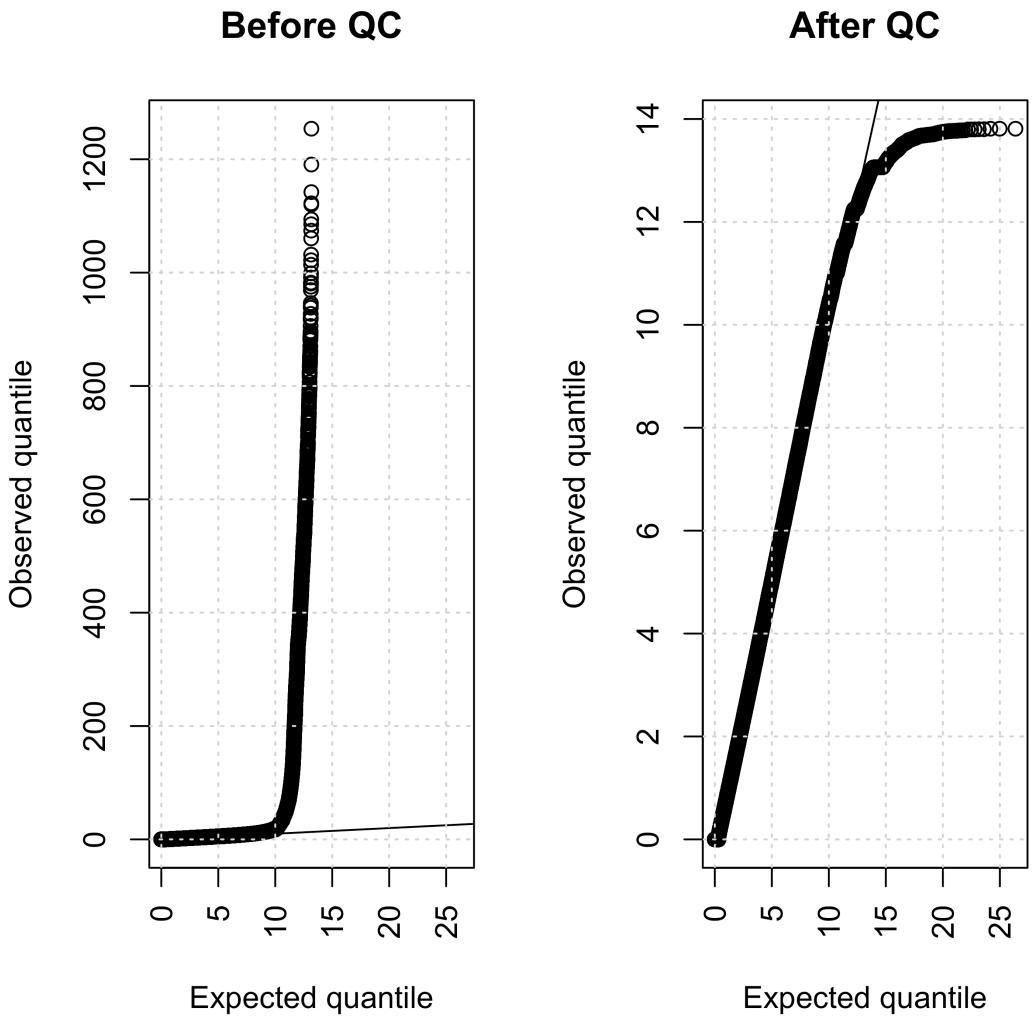


Table 2.1| A summary of QC filters applied on the PoBI data.

QC step	**QC Filter**	**Samples / SNPs before QC**	**Samples / SNPs removed**	**Samples / SNPs after QC**
Sex check	F < 0.2 = Female F > 0.8 = Male	3,702/547,862	173/0	3,529/547,862
Missingness rate	SNPs < 99% Samples < 99%	3,529/547,862	97/10,361	3,432/537,501
HWE	0.001	3,432/537,501	0/5,413	3,432/532,088
Cryptic relatedness	IBD >= 0.125	3,432/532,088	222/0	3,210/532,088
Population structure	-	3,210 /532,088	61/0	3,149/532,088

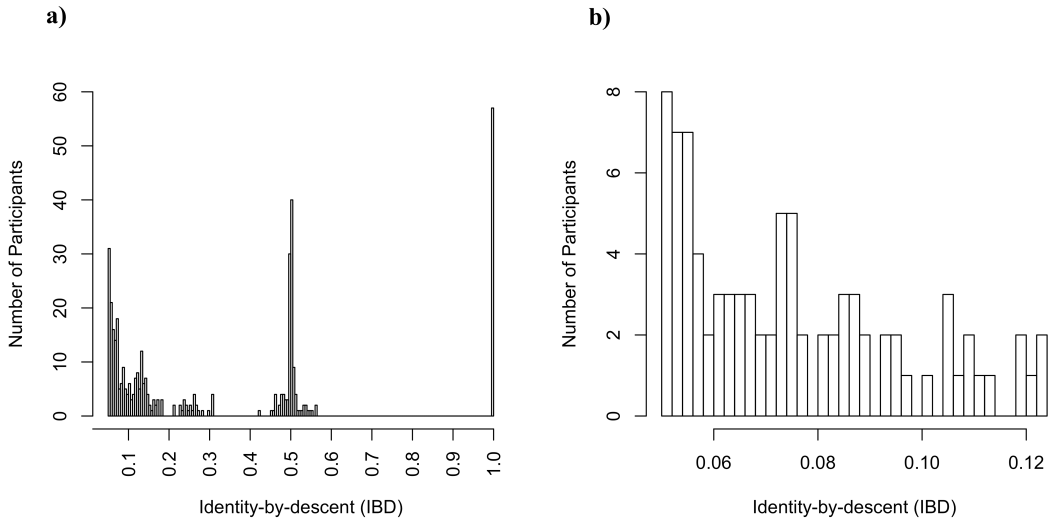
QC, quality control; SNP, Single nucleotide polymorphism; F, inbreeding coefficient; HWE, Hardy-Weinberg equilibrium; IBD, Identity-by-Descent

2.1.4 Cryptic relatedness

The statistical methods used to perform population-based association analyses such as GWASs rely on the assumption of having independent observations. For example, performing Fisher's exact test on population data with related individuals could inflate or deflate the test statistics, producing type I or type II errors. Cryptic relatedness (unknown common ancestry among smaller groups) and population structure (remote common ancestry among large groups of individuals) are among the factors that could result in the violation of the statistical assumptions (Astle & Balding, 2009; Devlin & Roeder, 1999).

One of the simplest ways to check for cryptic relatedness in genotype data is to estimate pairwise kinship coefficients among the individuals, in the form of pairwise allele sharing by Identity-by-descent (IBD) between individuals (Weale, 2010). For example, parent-child pairs on average share one allele IBD at each locus, while monozygotic twins always share two alleles IBD at each locus (Turner et al., 2011). Before calculating IBD in a GWAS dataset, it is advised to prune the data for LD as estimating IBD using the whole GWAS data could be a time-consuming process (Weale, 2010). After removing high LD SNPs (n = 408,243, $r^2 = 0.2$) from the current data, 123,845 remaining SNPs with low LD among them were used to calculate IBD between all individuals. It is advised to exclude individuals who share IBD at a proportion greater than 0.185 (Anderson et al., 2010) with at least one other individual, though a more stringent criterion was used in the current analysis (IBD $\geq$ 0.125). A total of 222 samples were excluded based on relatedness (Figure 2.4 and Table 2.1).

Figure 2.4 | Histogram showing the distribution of estimated Identity-by-descent (IBD) from the genotype data a) before excluding related samples (only relationships with IBD > 0.05 are shown here, n=407) and b) after excluding related samples. SNPs on non-autosomes were excluded from IBD calculation. An IBD value of 1, 0.5 and 0.25 describes duplicates/monozygotic twins, first degree and second degree relatives, respectively. An IBD value of ≥ 0.125 was used to exclude cryptic relatedness.



2.1.5 Population stratification and batch effects

When a particular ancestry group is over-represented in cases or controls, it may result in spurious genetic associations. This is commonly known as population structure or stratification (Anderson et al., 2010). As the cases and controls were recruited from a well-characterised population of the British Isles in the current study, population structure should not be a major issue in the PoBI data. However, it is important to check that the population structure matches our expectations (Leslie et al., 2015), and also to inspect the data for the effects of genotyping samples in different batches. Principal component analysis (PCA) has been widely used to correct for population structure in GWAS (Patterson et al., 2006), and can be used to identify geographic variations in genotypes.

Figure 2.5 and Figure 2.6 shows the largest three eigenvectors from the PCA performed on the PoBI dataset before and after outlier removal, respectively. Figure 2.7 shows the largest three PCs after performing outlier removal. SNPs with low minor allele frequency (MAF<.05) were excluded to produce this plot (Figure 2.7).

After performing quality-control on genotype and phenotype data, there were 1,593 samples left for the final analysis. Figure 2.8 and Table 2.2 show the geographical distribution of the samples used for the final analysis.

Figure 2.5 | Principal Component Analysis (PCA) of 3,210 individuals before performing outlier removal. High LD SNPs (r^2 = 0.2) were excluded before performing PCA. The colour-coded circles represent samples from different counties from the British Isles. The grey dots represent samples that do not have four grand-parents from the same area.

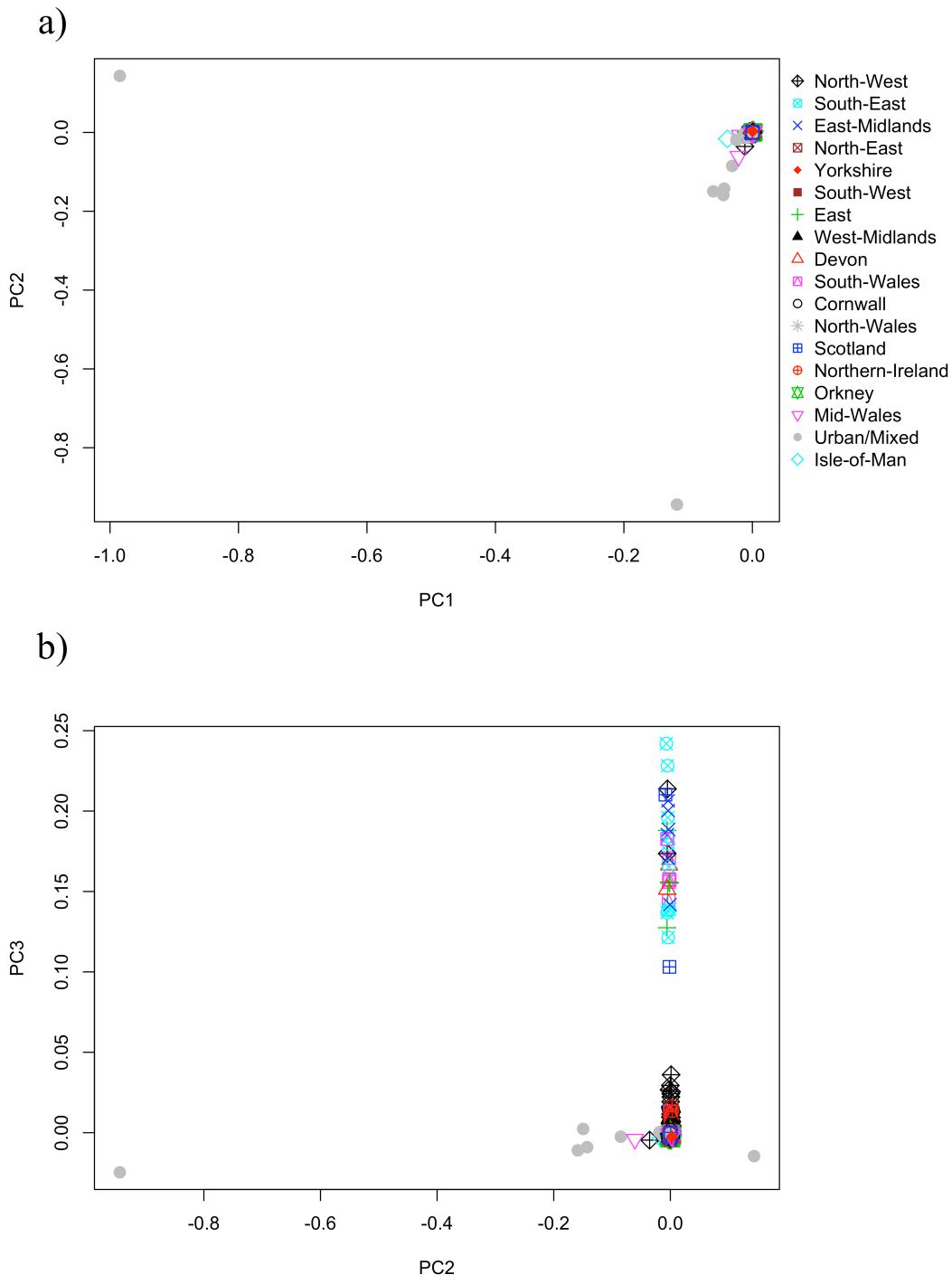


Figure 2.6 | Principal Component Analysis after outlier removal (n=3,149). The black-coloured diamonds represent samples from the county Cumbria, and may represent an extended family.

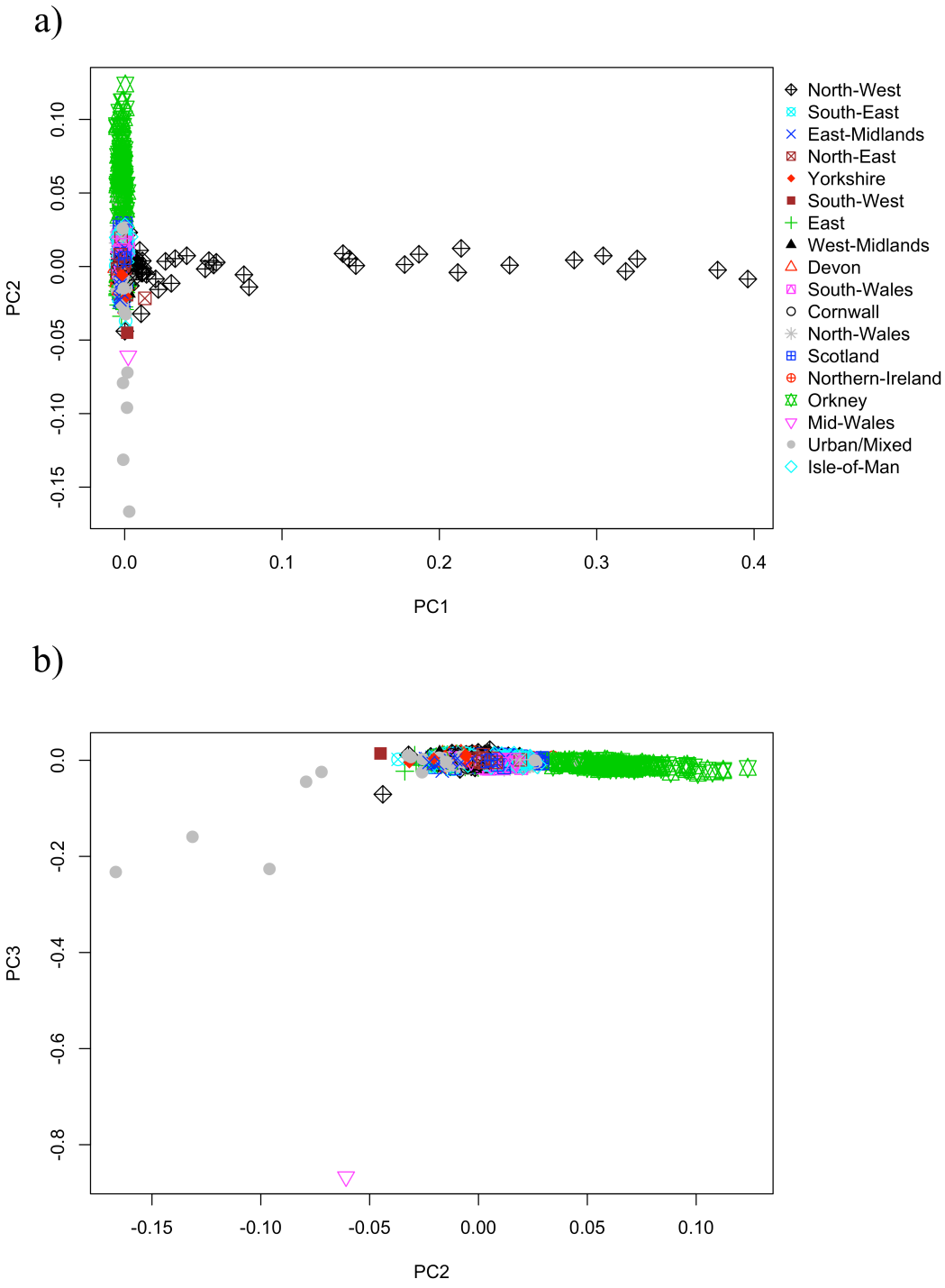


Figure 2.7 | Principal Component Analysis (PCA) of 3,149 individuals after performing outlier removal. SNPs with low MAF (<0.05) were excluded to produce this plot. The results are very similar to those reported in (Leslie et al., 2015).

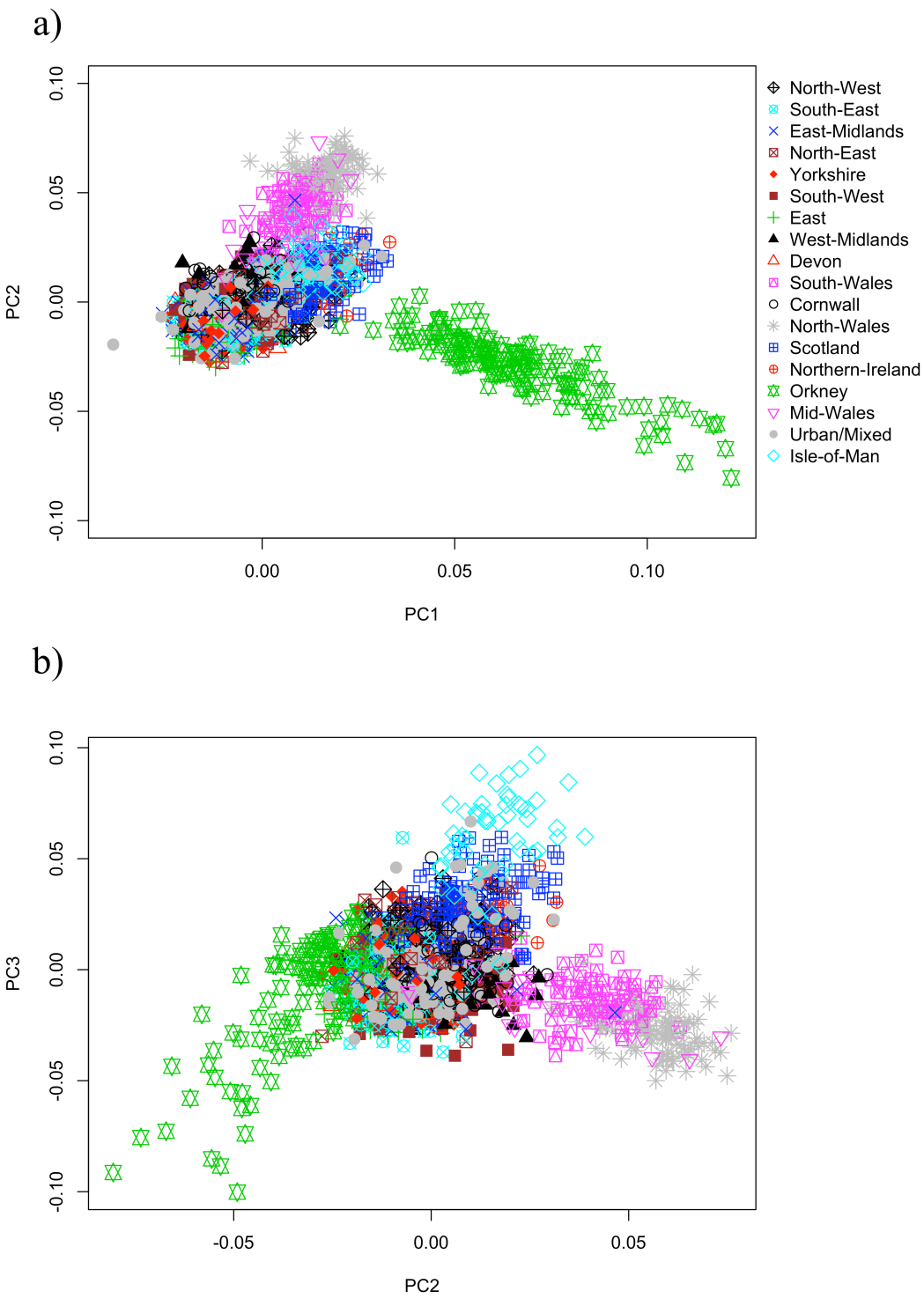
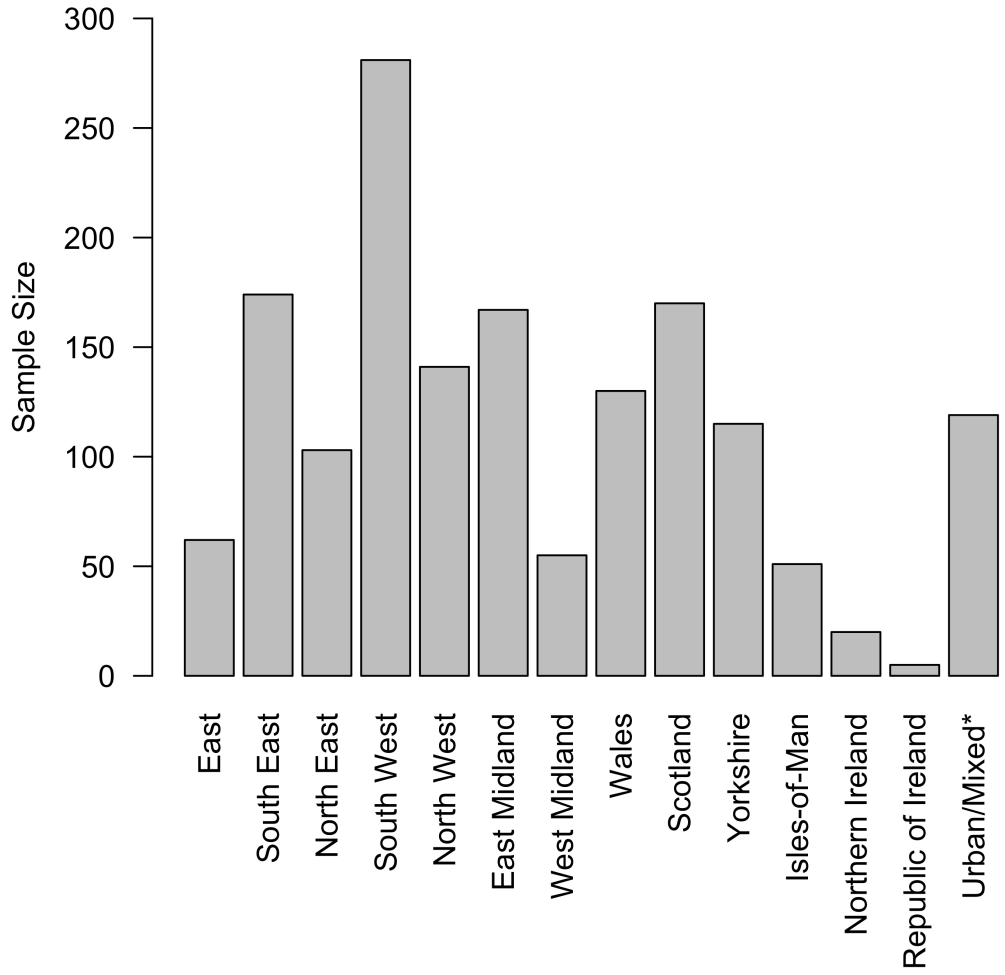


Figure 2.8 | A histogram showing the geographical distribution of samples used in the current analysis (Although 3,149 samples passed the quality control, only 1593, shown here, had both phenotype and genotype data available). Table 2.2 shows per-county county sample size.



*Urban/Mixed represents samples who did not fall into the criterion of having four grand-parents from the same area.

Table 2.2 | A per-county list of samples used in the current analysis (n=1,593).

Region	***County***	***Sample size***
East (n = 62)	Suffolk	9
	Norfolk	51
	Essex	2
South East (n = 174)	Oxfordshire	62
	Kent	46
	Surrey	1
	Hampshire	21
	Buckinghamshire	1
	Sussex	43
South West (n = 281)	Somerset	5
	Devon	19
	Gloucestershire	22
	Forest of Dean	33
	Wiltshire	1
	Dorset	26
	Cornwall	175
East Midland (n = 167)	Lincolnshire	73
	Leicestershire	43
	Nottinghamshire	18
	Northamptonshire	28
	Derbyshire	5
West Midland (n = 55)	Herefordshire	22
	Shropshire	1
	Warwickshire	2
	Worcestershire	24
	Staffordshire	6
North West (n=141)	Cheshire	9
	Cumbria	107
	Lancashire	25
Wales (n=130)	Mid Wales	27
	North Wales	55
	North Pembrokeshire	24
	Wales	8
	South Pembrokeshire	16
Scotland (n = 170)	Scotland	24
	Argyle	29
	Banff and Buchan	47
	Orkney	70
Northern Ireland		20
Republic of Ireland		5
Isles-of-Man		51
Yorkshire		115
North East		103
Urban/Mixed		119

Although 3,149 samples passed the quality control, only 1593, shown here, had both phenotype and genotype data available yet.

2.2 Discussion

Implementing stringent QC filters on GWAS data is important as poor QC filters could lead to the violation of assumptions made by statistical methods, resulting in type I or type II errors. Although QC procedures vary from study-to-study, a standard QC should be performed to remove sex inconsistencies, genotyping errors and population structure. In the current data, although the samples were collected from a well-characterised population of the British Isles using the criterion of having four grandparents from the same rural area, stringent QC filters were performed before doing association analysis. Based on different QC filters, a total of 561 samples and 15,774 SNPs were excluded before performing association analysis (Table 2.1). It is important to note that a large proportion of the samples excluded are due to relatedness, which is probably an unavoidable aspect of sampling locally in rural areas.

2.3 Conclusions

Adequate quality-control filters were applied to the PoBI data. In total, 561 samples and 15,774 SNPs were excluded based on QC filters such as sex inconsistencies, genotyping error, cryptic relatedness and population stratification. Table 2.1 shows a summary of the quality control filters applied to the PoBI data.

3 Population genetics of human pigmentation

3.1 Introduction

Pigmentation, one of the most studied visible phenotypes within and between human populations, is of key importance in the understanding of the genetics of complex traits. Pigmentation has a major role in protection against ultraviolet (UV) radiation (Jablonski & Chaplin, 2010) and melanoma cancer- a cancer of melanocytes, common in European populations (Rees, 2003, 2004). Hair, skin and eye colour are among the visible pigmentation phenotypes that vary to a great extent in human populations, particularly among Europeans (Wilde et al., 2014). This variation in pigmentation phenotypes is due to different biochemical reactions that synthesise melanin- a pigment synthesised in melanocytes. Melanin, divided into two categories: Eumelanin (subtypes Brown/Black) and Pheomelanin (subtypes Yellow/Red), characterises individuals' hair, skin and eye colour (Nan et al., 2009). Variation in these visible traits is more common in individuals of European ancestry, while some other populations (e.g. Asians) are essentially fixed for black hair and brown eyes (Sulem et al., 2007). Beside conferring protection from UV radiation, pigmentation is also of interest due to being a likely target of sexual selection. It was proposed that sexual selection – the preference for light skin females among males, caused human pigmentation diversity, not natural selection (Darwin, 1871; Ihara & Aoki, 1999). However, previous studies (using data sets derived from measurements with standard

reflectance spectrophotometric devices) do not support the hypothesis that sexual selection has shaped the skin colour variation in humans (Madrigal & Kelly, 2007).

The main aim of the pigmentation research performed in the last decade has been; 1) to understand the molecular basis of human pigmentation; 2) to understand how gene-gene (GXG) interaction determines these complex traits and 3) how natural selection, or evolution more generally, has shaped these genes. Different studies, including functional expression, candidate gene approaches, sequence based analysis and genome-wide association studies (GWAS) have shaped our knowledge of these complex traits. Genetic variants identified through these approaches, including those in *MC1R, OCA2, IRF4, ASIP*, and *SLC45A2* and a newly identified variant in *KITLG* (Guenther et al., 2014; Sulem et al., 2007), have fundamental roles in determining this complex trait. The proteins encoded by these genes determine the proportion of eumelanin relative to pheomelanin, in turn affecting pigmentation.

Understanding the biological functions of these variants has also helped us in understanding the molecular basis of melanoma cancer and also in the prediction of human phenotypes (such as eye and hair colour) in forensic science which may aid in the identification of suspects from DNA left at crime scenes (Branicki et al., 2011; Maronas et al., 2015). For example, there are DNA tools, such as IrisPlex and HIrisPlex, which can be used for accurate prediction of externally visible characteristics based on carefully chosen genetic markers (Walsh et al., 2011). IrisPlex can be used for eye colour prediction based on six markers while HIrisPlex can predict both, hair and eye colour, based on twenty-four markers (Walsh et al., 2013).

3.2 GWAS of hair pigmentation

3.2.1 Introduction

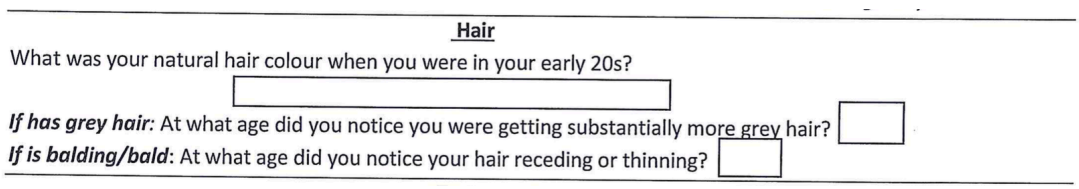
Hair colour in human populations is a complex trait, which is highly heritable, and it has been reported that a large proportion of variation (61-92%) in natural hair colour can be explained by genetic factors in Europeans while remaining variation is due to environmental factors or measurement errors (Brauer & Chopra, 1978; Lin et al., 2015; Sulem et al., 2007). With the advancement in genotyping technologies, it is now feasible to scan the human genome for genetic factors underlying complex traits such as hair and skin colour. The genetics underlying different hair colours such as red, black and blond has been known for almost a decade now with some genome-wide studies claiming new insights into the hair pigmentation trait (Gerstenblith et al., 2010).

A handful of genome-wide association studies have been performed to identify genes underlying hair colours. For example, a GWAS published in 2007 found a new variant *KITLG* that is significantly associated with blond hair in individuals of northern European ancestry (Sulem et al., 2007). Another GWAS (Han et al., 2008) has shown that *IRF4* (interferon regulatory factor 4) and *SLC24A4* (solute carrier family 24 member 4) are linked to human hair and skin pigmentation. These findings, therefore, raise the question whether there remain any unidentified genetic variants that underlie pigmentation phenotypes and whether studies with new approaches, e.g. using a well-characterised population, would be able to provide new insights in this area of research. The present research uses genome-wide association analysis to study hair colour differences within the people of the British Isles.

3.2.2 Phenotype data collection

Subjects: Participants were recruited from different counties of the British Isles using the criterion of having four grandparents from the same area. For hair colour, each participant was asked to report their natural hair colour when they were in their early 20's (Figure 3.1). Although hair colour is a quantitative trait and categorization of the trait could result in loss of power (Liu et al., 2010), most of the GWAS performed to identify variants that underlie hair colour have been done using categorical data (Han et al., 2008; Lin et al., 2015; Sulem et al., 2007). For phenotypes, less amenable to categorical description, such as skin colour, quantitative measurements were taken. For hair greying phenotype, participants were asked – at what age did they notice hair greying? Figure 3.1 shows the questions asked to collect the phenotype data (Appendix A.2). It is quite apparent from Figure 3.3 that most subjects noticed substantial hair greying around the age of 50 which, to some extent, fits in the 50/50/50 rule of thumb ('at age 50 years, 50% of the population has at least 50% grey hair') about hair greying (Keogh & Walsh, 1965). For analysis, participants were divided into four major categories of hair colour: Brown (n = 1,127, includes light, medium and dark-brown), Blond (n = 254, includes light, medium, dark-blond as well as fair), Red (n = 91, includes auburn and light-red), Black (n = 113 including dark hair) as shown in Table 3.1. Figure 3.2 and Figure 3.3 show the distribution of phenotypes collected (n = 1,585 for hair colour and n = 1,416 for hair greying).

Figure 3.1 | A part of the questionnaire form used for hair colour data collection.


Hair

What was your natural hair colour when you were in your early 20s?

If has grey hair: At what age did you notice you were getting substantially more grey hair?

If is balding/bald: At what age did you notice your hair receding or thinning?

Table 3.1 | Characteristics of participants in the hair colour GWAS.

Hair Colour	***N (Male/Female)***	***Mean age (SD)***
Brown	1127 (528/599)	65.55 (11.75)
Blond	254 (109/145)	64.87(11.78)
Red	91 (39/52)	64.58(10.79)
Black	113 (73/40)	68.32(10.67)

Figure 3.2 | Distribution of hair colour data (n = 1,585).

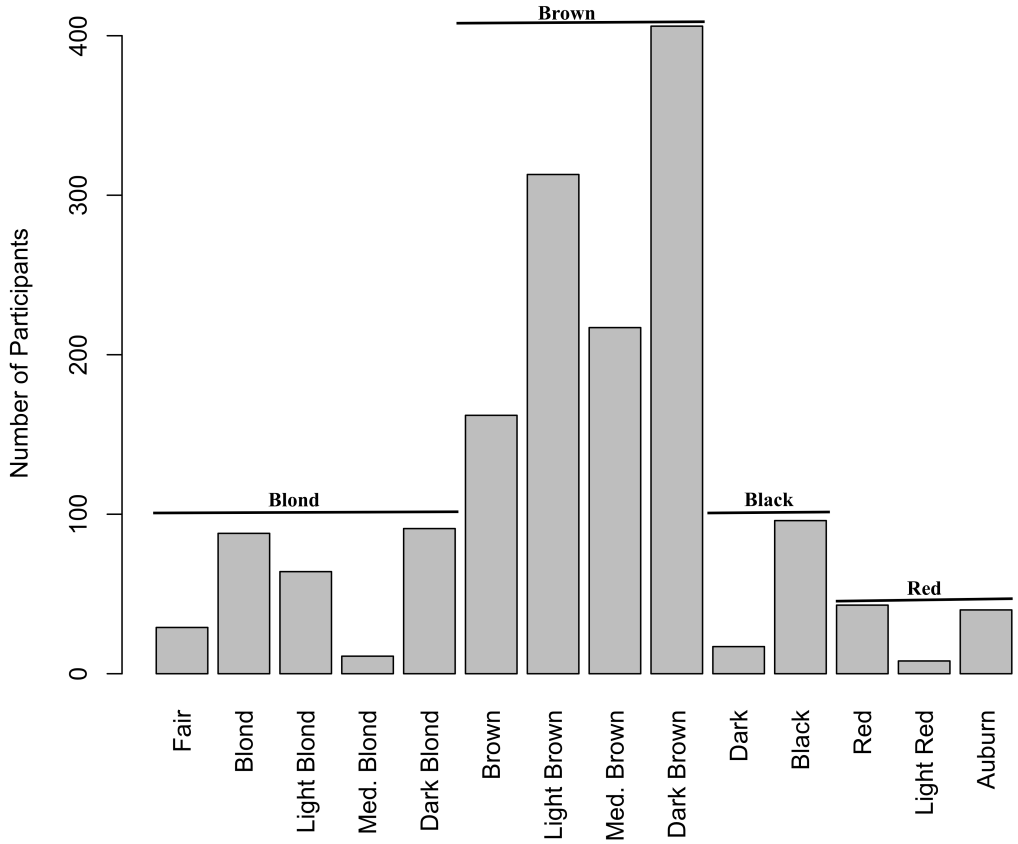
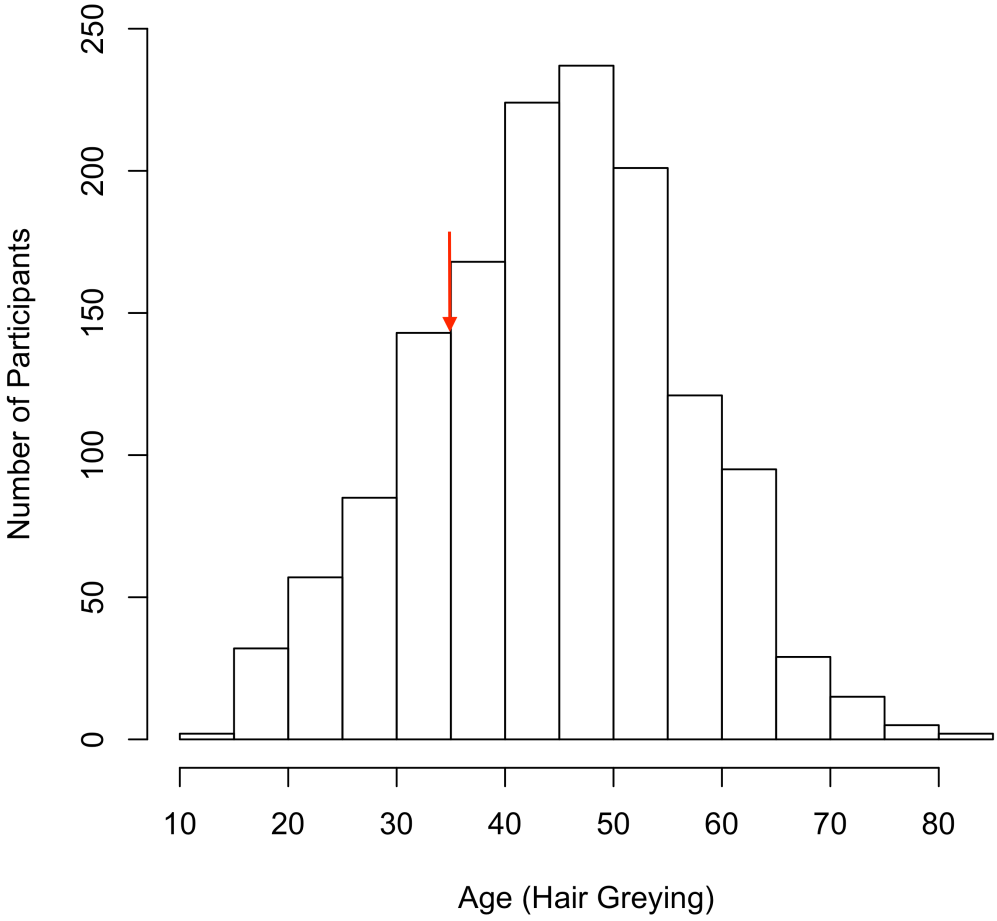


Figure 3.3 | Distribution of hair greying (age) phenotype data (n = 1,416). The red line indicates the cut off chosen for association analysis (In order to investigate the genetic variants conferring large effects, subjects who noticed early greying before the age of 35 (lower extremes) were used as cases, in this case ~15%).



3.2.3 Statistical analysis

After performing quality control on genotype data, a total of 1,585 (n = 749 male, n = 836 female, mean age = 65.6 years, SD = 11.6) subjects were considered for further analysis. Based on the previous studies (Lin et al., 2015; Sulem et al., 2007), hair

colour phenotype data were subjected to four comparisons: 1) Red versus non-red; 2) Blond versus non-blond; 3) Black versus non-black and 4) Light (blond and red) versus dark (brown and black).

3.2.4 Red hair

Association analysis for red hair was performed in a total of 1,585 subjects using the Fisher's exact test (3×2 genotypic test). Standard QC procedures on genotype data were applied before doing an association analysis. For calculating odds ratios 0.5 was added to each cell of the genotypes (Haldane, 1956). Exclusion of SNPs based on the minor allele frequency (MAF < 0.01) did not produce any difference in the association statistics that were significant (results not shown); therefore, no SNPs were removed via a MAF threshold for this particular trait. A total of twenty-three single nucleotide polymorphisms (SNPs) reached the genome-wide significance threshold ($P \leq 5 \times 10^{-8}$, Table 3.2, Figure 3.4, Appendix B.1 for the list of all the SNPs reached genome-wide significance level). In addition to the GWAS hits, SNPs which attained suggestive genome-wide significance ($P \leq 1\times10^{-5}$) in the current research have also been discussed where necessary. The volcano plot (Figure 3.5) shows the observed p-values and odds ratios for the SNPs analysed in the data. A sharp deviation of observed p-values from those expected under the null hypothesis of no association in the Q-Q plot (Appendix B.2) supports the peak observed in the Manhattan plot (Figure 3.4).

All SNPs passing genome-wide threshold were within 1 Mb region on Chromosome 16, containing the well-known pigmentation gene *MC1R* (Figure 3.6). The most significant SNP was rs12918773 ($P = 1.0\times10^{-26}$, $OR_{Dom} = 8.17$, Table 3.2 and Table 3.3) on chromosome 16 (16q24.3) is ~ 0.2 Mb from the *MC1R* gene and it was shown to be associated with facial pigmentation spots (i.e. freckles) in a GWAS of the Dutch

subjects (Jacobs et al., 2015). rs258322 ($P = 1.2\times10^{-24}$, $OR_{Dom} = 7.72$, $r^2 = 0.09$, Appendix B.3) was the second most significant association with red hair. rs258322 is an intron variant of *CDK10* (a cell-cycle regulator) gene which had been shown to affect melanoma risk and human pigmentation in a previous study (Bishop et al., 2009). Previous studies have also shown that rs258322 is in linkage disequilibrium with an *MC1R* allele at rs1805007 (Antonopoulou et al., 2015). Given that rs258322 (*CDK10*) is in strong LD with rs12918773 (Table 3.2), it is more likely that the top hit is tagging *MC1R*.

Table 3.2 | Association of genetic variants with red versus non-red hair colour in PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***p-value[b]***	***OR***	***r^2***
rs12918773	16	89741403	*SPATA33/MC1R*	1.0×10^{-26}	8.17	-
rs258322	16	89755903	*CDK10/MC1R*	1.2×10^{-24}	7.72	0.90
rs7195043	16	90020861	*DEF8*	8.3×10^{-20}	8.04	0.05
rs4785751	16	90029417	*DEF8*	3.1×10^{-18}	7.16	0.05
rs9850626	3	126068779	*KLF15*	2.5×10^{-7}	4.90	-

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with top SNP rs12918773); [b]Fisher's exact test; OR, Odds ratio

Table 3.3 | Distribution of rs12918773 genotypes in participants with red and non-red hair. Shown here is the test of association (genotypic) in a 2 × 3 table of phenotype-by-genotype. *AA* and *aa* represent minor and major allele, respectively.

Genotype	***AA***	***Aa***	***aa***	***Total***
Red	15	46	30	91
Non-red	8	289	1196	1493
Total	23	335	1226	1584

$P_{Dom} = 8.9 \times 10^{-21}$, $OR_{Dom} = 8.17$; $P_{Rec} = 2.6 \times 10^{-14}$, $OR_{Rec} = 36.35$

Figure 3.4 | Manhattan plot for the GWAS of red versus non-red hair. The observed – log_{10} transformed p-values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). The horizontal red and blue lines indicate the threshold for genome wide significance ($P = 5 \times 10^{-8}$) and suggestive genome-wide significance ($P = 1 \times 10^{-5}$), respectively.

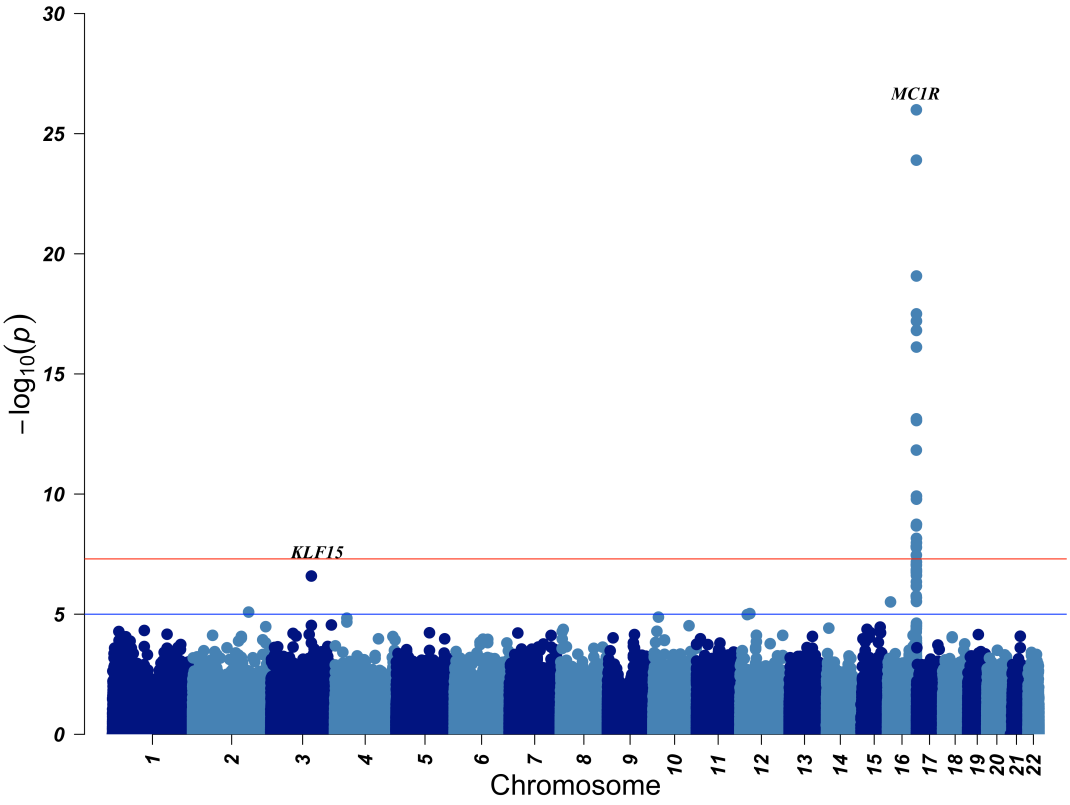
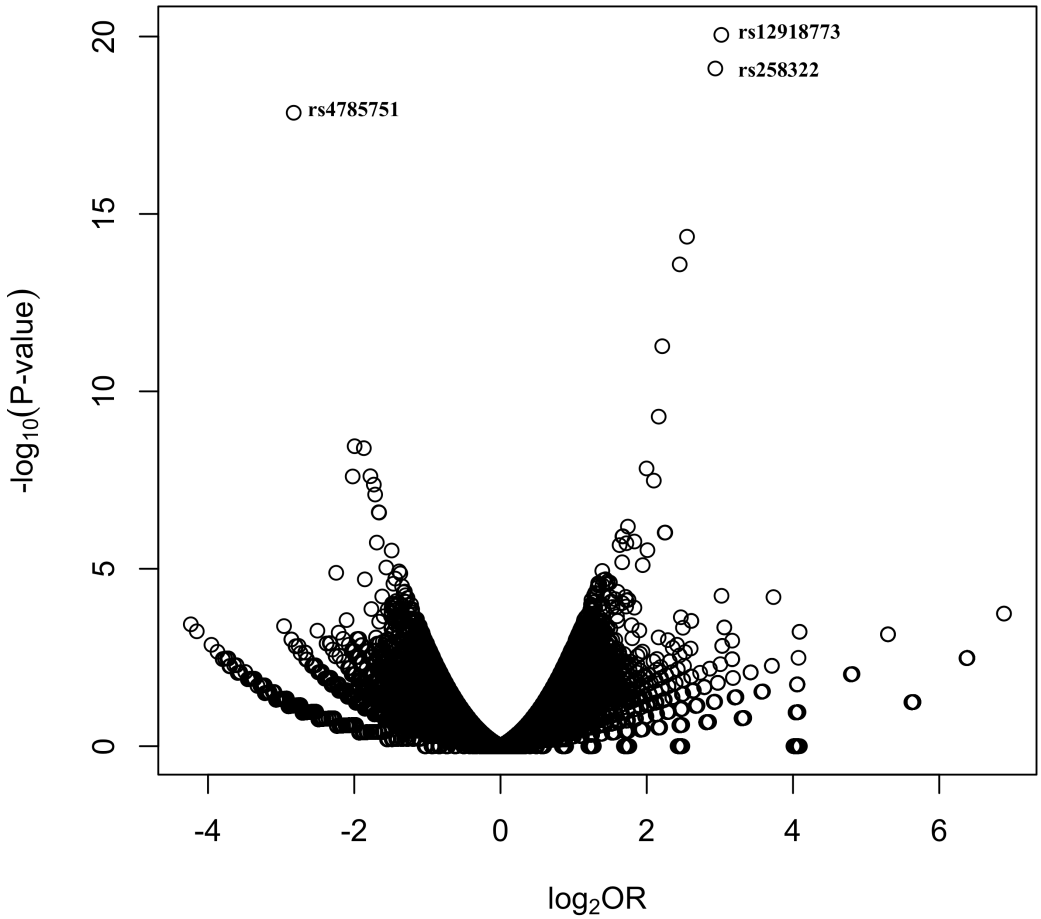


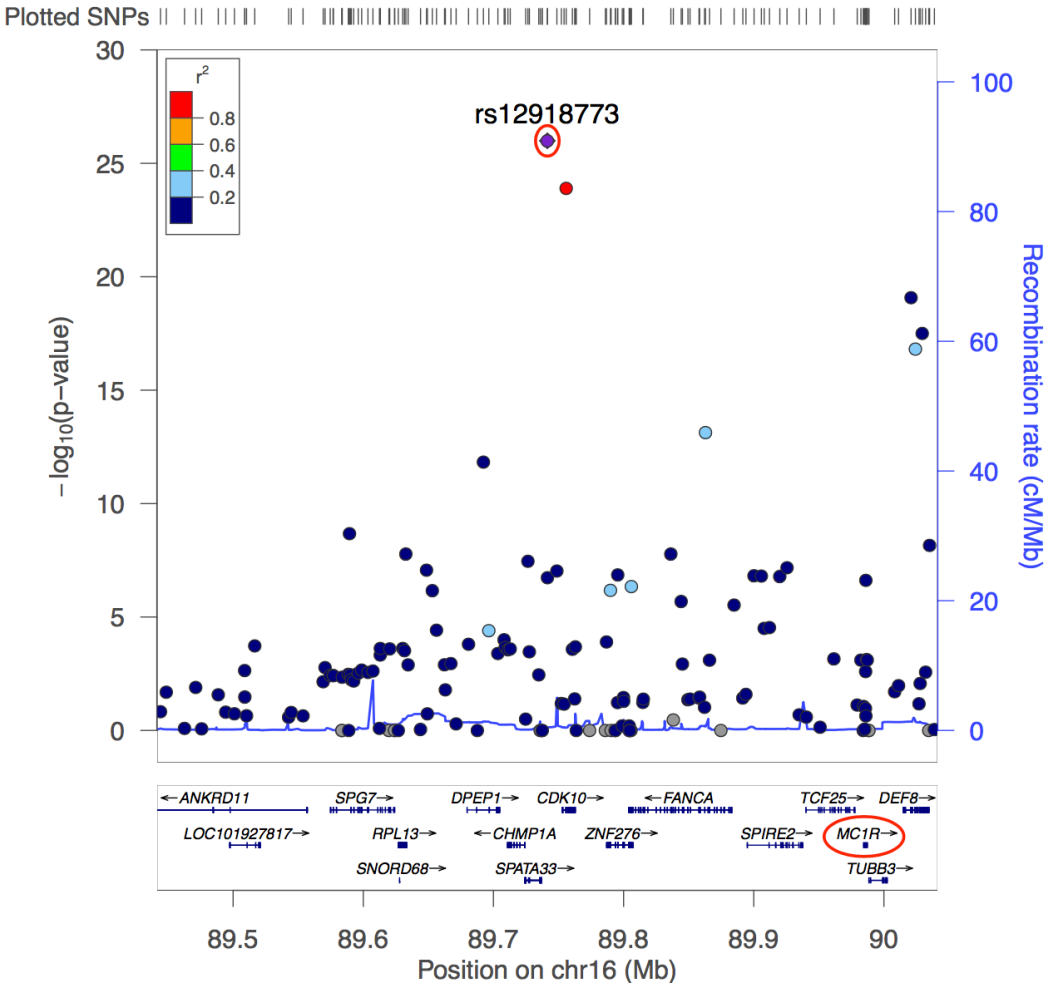
Figure 3.5 | Volcano plot for the GWAS of red versus non-red hair. The observed – log_{10} transformed p-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).



In addition to the significant hits on chromosome 16, no other SNPs reached genome-wide significance level ($P \leq 5 \times 10^{-8}$) but rs9850626 ($P = 2.5 \times 10^{-7}$, $OR_{Dom} = 4.90$, *KLF15,* Appendix B.4) on chromosome 3 attained suggestive genome-wide significance ($P = 1 \times 10^{-5}$). *KLF15* is a protein coding gene and a member of the krüppel-like factor family, a family of transcription factors with zinc finger domains, which plays a key role in cell differentiation and regulation of the circadian rhythm (Yoshitane et al., 2014). A study has demonstrated that *KLF15* interacts with peripheral clock components *BMAL1/CLOCK* to regulate cardiac repolarization and arrhythmogenesis in mice (Jeyaraj, Haldar, et al., 2012). In addition, it is interesting to note that *BMAL1* (a core clock gene) has been shown to regulate human pigmentation in hair follicles and knockdown experiments in the hair follicles have been shown to increase melanosome biogenesis in hair follicle melanocytes (Hardman et al., 2015). Beside its interaction with *BMAL1*, *KLF15* has been shown to regulate *TGF-β* in mouse heart (Yu et al., 2015). *TGF-β* has also been shown to downregulate melanogenesis associated transcription factor *(MITF)* in the melanocyte stem cells in hair follicles (Nishimura et al., 2010). Taken together, although not directly, the above-mentioned evidence suggests that *KLF15* may have a role in human pigmentation and further research is needed to unravel biological functions of *KLF15* in the lumens of hair follicles.

Other SNPs which attained suggestive genome-wide significance were rs10153883 ($P = 8.1 \times 10^{-6}$, $OR_{Rec} = 5.84$) on chromosome 2 and rs17692219 ($P = 9.4 \times 10^{-6}$, $OR_{Rec} = 9.45$) on chromosome 12.

Figure 3.6 | Regional association plot of the *MC1R* region. Observed $-\log_{10}$ values were plotted against the genomic location on chromosome 16 (q24.3). The purple diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes in a particular region are shown in the bottom.



3.2.5 Black versus non-black hair

Testing 113 black-haired subjects against 1,472 non-black-haired subjects using Fisher's exact test (3 × 2 genotypic test) showed six genome-wide significant hits on chromosomes 15 (four SNPs in *HERC2*), 6 (one SNP in *IRF4*) and 5 (one SNP in *SLC45A2*), replicating the known associations in pigmentation genes *HERC2*, *IRF4* and *SLC45A2* (Figure 3.7, Table 3.4). Among all four SNPs reaching genome-wide significance on chromosome 15, rs12913832 ($P = 6.7 \times 10^{-11}$, $OR_{Dom} = 3.61$ for minor allele positive association with having black hair, Table 3.4 and Table 3.5) showed the highest degree of association with black hair. All four SNPs on chromosome 15 were within 300 kb of *HERC2* (Appendix B.5). Table 3.4 shows a summary of the most significant SNPs and their linkage disequilibria (LD, r^2) in the *HERC2* region. rs3935591 and rs916977 showed strong *LD* with the most significant SNP (rs12913832). The volcano plot is shown in Figure 3.8. A clear deviation only in the tail of the Q-Q plot (Appendix B.2) indicates a true association and supports observations of the Manhattan plot (Figure 3.7).

A SNP on chromosome 6, rs12203592 in *IRF4* has been reported to affect human pigmentation traits (Han et al., 2008). rs12203592 also showed an association with hair greying in a GWAS in admixed Latin Americans (Adhikari et al., 2016). rs12203592 ($P = 3.3 \times 10^{-10}$, $OR_{Dom} = 3.67$ minor allele positively associated with non-black hair, Appendix B.6) was genome-wide significant in the current analysis of black versus non-black hair.

Solute carrier family 45 member 2 (*SLC45A2*) is a membrane-associated transporter protein (MATP) that mediates melanin synthesis. A SNP (rs35391, $P = 4.5 \times 10^{-8}$, $OR_{Dom} = 6.43$ minor allele positively associated with non-black hair, Appendix B.7)

in *SLC45A2* on chromosome 5 (p13.2) was significantly associated with the black- versus non-black hair in the PoBI data.

Table 3.4 | Association of genetic variants with black versus non-black hair in PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***P-value[b]***	***OR***	***r^2***
rs12913832	15	28365618	*HERC2*	6.7×10^{-11}	3.61	-
rs3935591	15	28374012	*HERC2*	1.1×10^{-9}	3.58	0.55
rs916977	15	28513364	*HERC2*	4.1×10^{-9}	3.33	0.59
rs12203592	6	396321	*IRF4*	3.3×10^{-10}	3.67	-
rs35391	5	33955673	*SLC45A2*	4.5×10^{-8}	6.43	-

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with top SNP rs12913832); [b]Fisher's exact test; OR, Odds ratio

Table 3.5 | Distribution of rs12913832 genotypes in participants with black and non-black hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Black	15	60	38	113
Non-black	60	459	952	1471
Total	75	519	990	1584

$P_{Dom} = 1.3\times10^{-10}$, $OR_{Dom} = 3.61$; $P_{Rec} = 0.0001504$, $OR_{Rec} = 3.59$

Figure 3.7 | Manhattan plot for the GWAS of black versus non-black hair. The observed $-log_{10}$ transformed p-values (Fisher's exact test) were plotted (y axis) against their chromosomal positions (x axis). The horizontal red line indicates the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

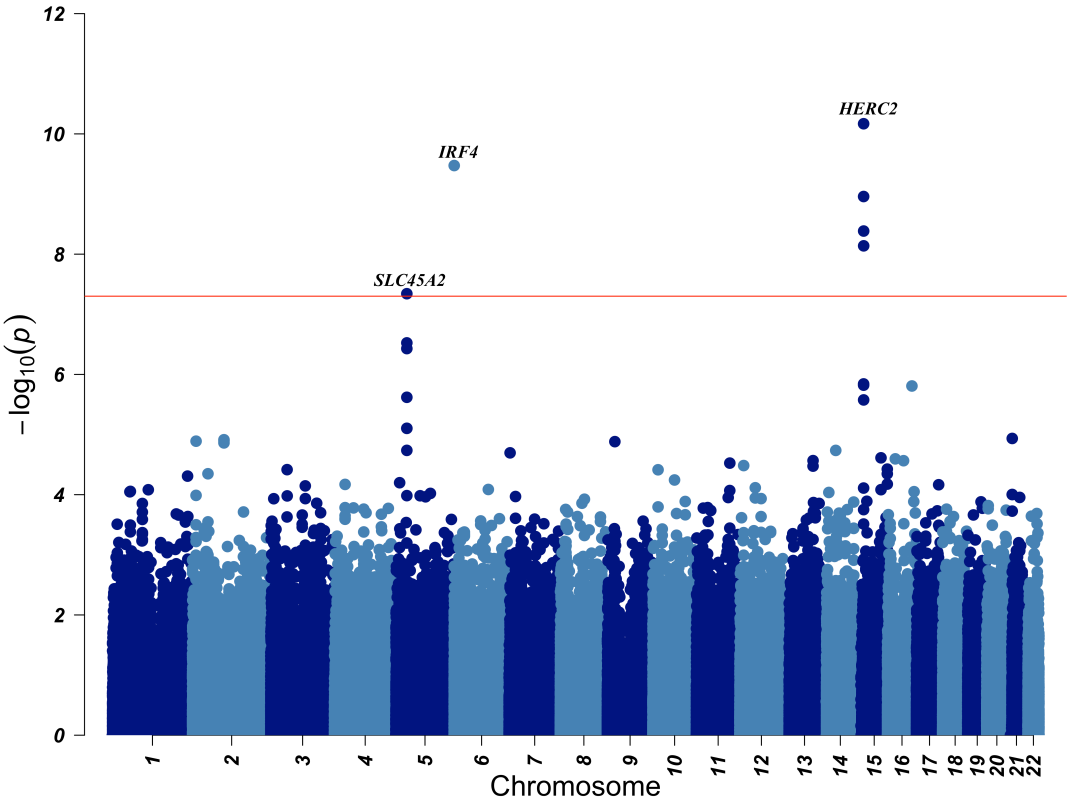
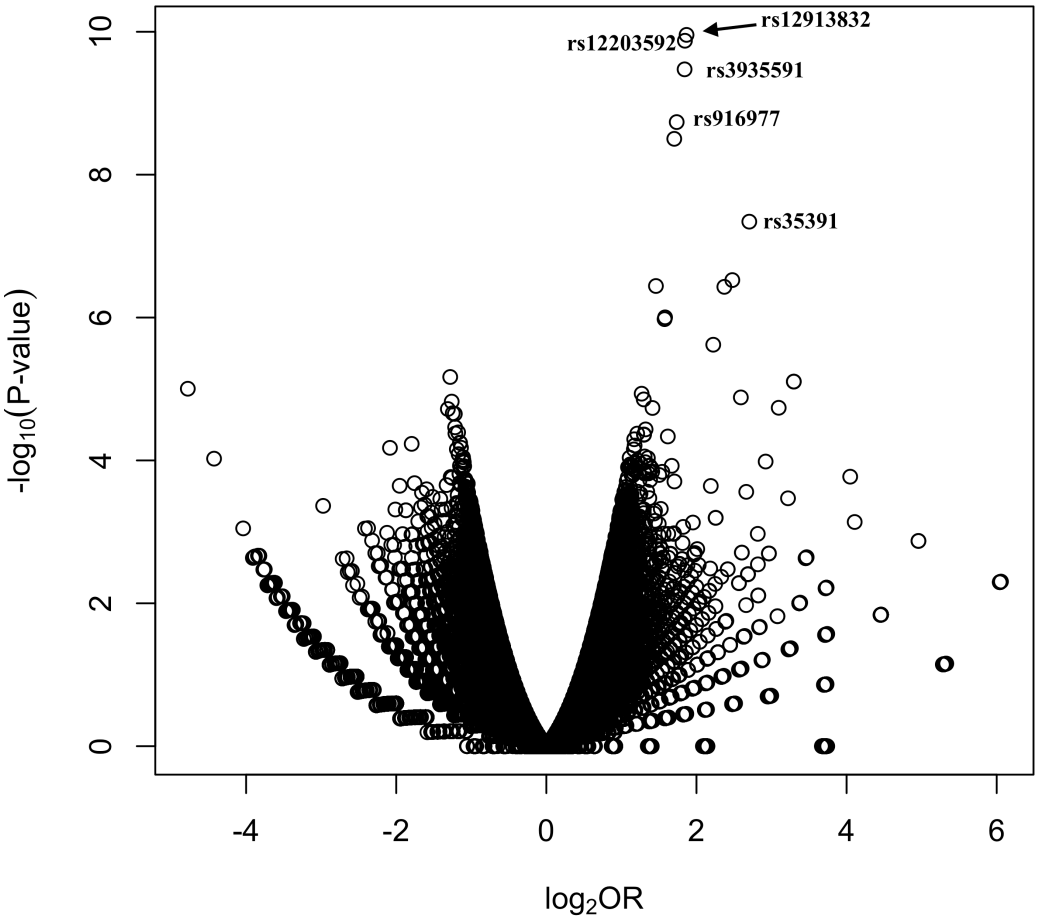


Figure 3.8 | Volcano plot for the GWAS of black versus non-black hair. The observed $-log_{10}$ transformed p-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).



3.2.6 Blond versus non-blond

254 subjects with blond hair were tested against 1,098 subjects with non-blond hair for genetic differences at 519,188 SNPs using the Fisher's exact test (3 x 2 genotypic test). Three pigmentation candidate genes, *IRF4, HERC2* and *KITLG,* collated from the literature (Lin et al., 2015; Sulem et al., 2007), had SNPs that were statistically significant ($P < 0.05$, given that *IRF4*, *HERC2* and *KITLG* are well known pigmentation genes). One SNP on chromosome 6 in *IRF4* and three SNPs on chromosome 15 in the *HERC2* region, reached genome-wide significance ($P \leq 5 \times 10^{-8}$, Figure 3.9). rs12203592 ($P = 1 \times 10^{-11}$, $OR_{Dom} = 3.15$ minor allele positively associated with non-blond hair) in *IRF4* was most strongly associated with blond versus non-blond hair (Table 3.6, Table 3.7 and Appendix B.5). Besides rs12203592, three SNPs in *HERC2* showed significant association with this phenotype (Table 3.6). All SNPs on chromosome 15 (Appendix B.8 for rs12913832 genotypes) were within 1 Mb region encompassing *HERC2*. A mutation in *HECT* and RLD domain containing E3 ubiquitin protein ligase 2 (*HERC2*) is known to affect the expression of the gene oculocutaneous albinism II (*OCA2*) which is associated with skin and eye pigmentation (Eiberg et al., 2008). The volcano plot and the Q-Q plot is shown in Figure 3.10 and Appendix B.2, respectively.

A GWAS published in 2007 (Sulem et al., 2007) showed that a variant near *KITLG* (encoding the ligand for *KIT* receptor tyrosine kinase) is associated with blond hair phenotypes in the European population. *KITLG* mediates development and maintenance of melanocytes in adult skin (Wehrle-Haller, 2003). Studies have also shown that rs12821256, a regulatory region variant in *KITLG*, is associated with blond hair phenotypes (Eriksson et al., 2010a; Han et al., 2008). This variant,

rs12821256, showed statistically significant association with blond versus non-blond status in the current data (P = 0.000128, OR_{Dom} = 1.76). Interestingly, in the current analysis, a different SNP (rs11610915, P = 9.7×10^{-5}, OR_{Dom} = 2.13, Table 3.8, Figure 3.11) in the *KITLG* region showed an even stronger association than rs12821256.

Table 3.6 | Association of genetic variants with blond versus non-blond hair colour in the PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***P-value[b]***	***OR***	***r^2***
rs12203592	6	396321	*IRF4*	1.0×10^{-11}	3.15	-
rs12913832	15	28365618	*HERC2*	1.1×10^{-11}	3.20	-
rs4778241	15	28338713	*HERC2*	1.6×10^{-8}	2.80	0.44
rs1667394	15	28530182	*HERC2*	3.6×10^{-8}	2.94	0.63

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with top SNP rs12913832); [b]Fisher's exact test; OR, Odds ratio

Table 3.7 | Distribution of rs12203592 genotypes in participants with blond and non-blond hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Blond	3	40	210	253
Non-blond	63	368	667	1098
Total	66	408	877	1351

$P_{Dom} = 3.6\times10^{-12}$, OR_{Dom} = 3.15; P_{Rec} = 0.001034, OR_{Rec} = 5.06

Table 3.8 | Association of genetic variants in *KITLG* with blond versus non-blond hair.

SNP	***Chr***	***Position***	***Gene***[a]	***p-value***[b]	***OR***	r^2
rs11610915	12	88902041	*KITLG*	9.7×10^{-5}	2.13	-
rs12821256	12	89328335	*KITLG*	0.000128	1.76	0.29

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with top SNP rs11610915); [b]Fisher's exact test; OR, Odds ratio

Figure 3.9 | Manhattan plot for the GWAS of blond versus non-blond hair. The observed $-log_{10}$ transformed p-values (Fisher's exact test) were plotted (y axis) against their chromosomal positions (x axis). The horizontal red line indicates the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

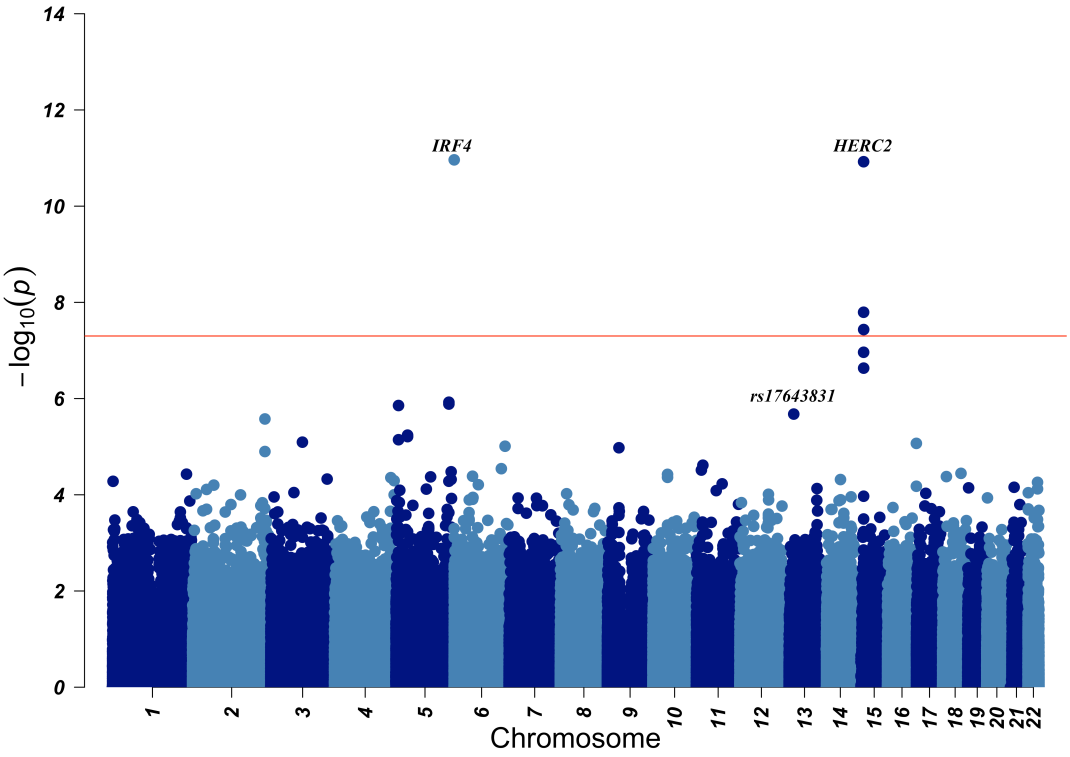


Figure 3.10 | Volcano plot for the GWAS of blond versus non-blond hair. The observed $-log_{10}$ transformed P-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).

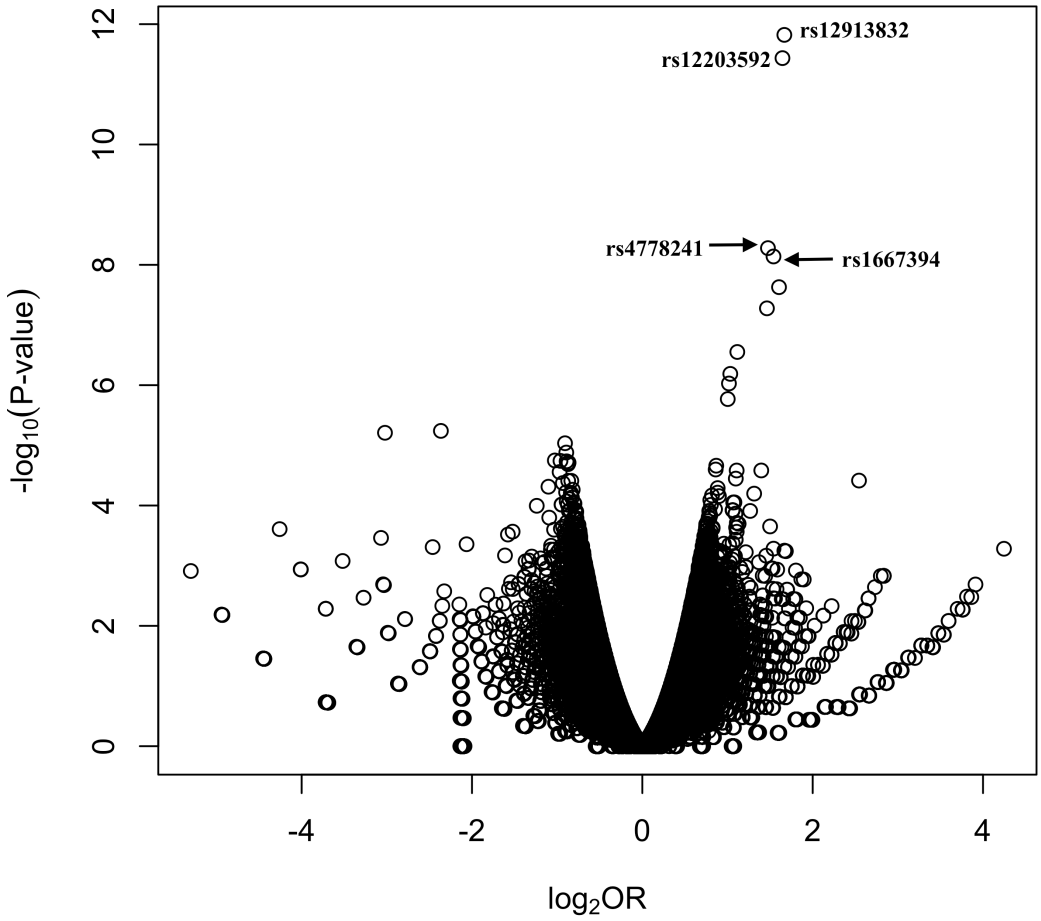
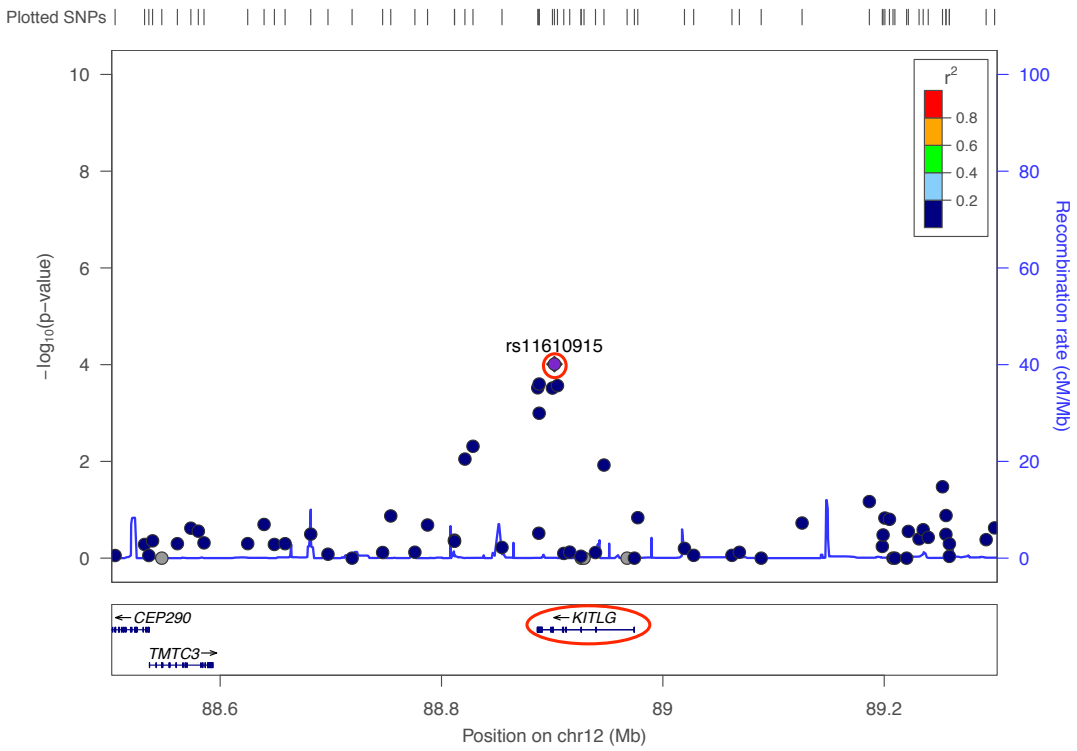


Figure 3.11 | Regional association plot showing the *KITLG* region on chromosome 12. Observed $-log_{10}$ values were plotted against the genomic location on chromosome 12 (q21.32). The purple diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.



3.2.7 Light versus dark hair

Comparing light-haired subjects (blond and red) versus dark-haired subjects (brown and black) also replicated the known genes associated with pigmentation on chromosomes 6, 15 and 16. The Manhattan plot (Figure 3.12) shows genome-wide hits on chromosome 16 (*MC1R* region), chromosome 15 (*HERC2*, Appendix B.9 for rs12913832 genotypes) and chromosome 6 (*IRF4*, Appendix B.10 for rs12203592 genotypes). Two SNPs (rs12918773, rs258322) on chromosome 16 within 300 kb of *MC1R* gene showed strongest association with light versus dark hair phenotype (Table 3.9 and Appendix B.11). Table 3.10 shows the genotype distribution for rs12918773. The volcano and the Q-Q plots are shown in Figure 3.13 and Appendix B.2, respectively.

Table 3.9 | Association of genetic variants with light versus dark hair in PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***P-value[b]***	***OR***	***r^2***
rs12918773	16	89741403	*SPATA33/MC1R*	5.4×10^{-20}	2.75	-
rs258322	16	89755903	*CDK10/MC1R*	6.4×10^{-20}	2.64	0.90
rs7195043	16	90020861	*DEF8*	4.7×10^{-11}	2.50	0.05
rs12913832	15	28365618	*HERC2*	4.6×10^{-11}	2.52	-
rs12203592	6	396321	*IRF4*	7.1×10^{-10}	2.36	-

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with SNP rs12918773); [b]Fisher's exact test; OR, Odds ratio

Table 3.10 | Distribution of rs12918773 genotypes in participants with light and dark hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Light hair	21	111	213	345
Dark hair	2	220	988	1210
Total	23	331	1201	1555

$P_{Dom} = 8.8\times10^{-14}$, $OR_{Dom} = 2.75$; $P_{Rec} = 1.8 \times 10^{-12}$, $OR_{Rec} = 39.04$

Figure 3.12 | Manhattan plot for the GWAS of light versus dark hair. The observed $-log_{10}$ transformed p-values (Fisher's exact test) were plotted (y axis) against their chromosomal positions (x axis). The horizontal red line indicates the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

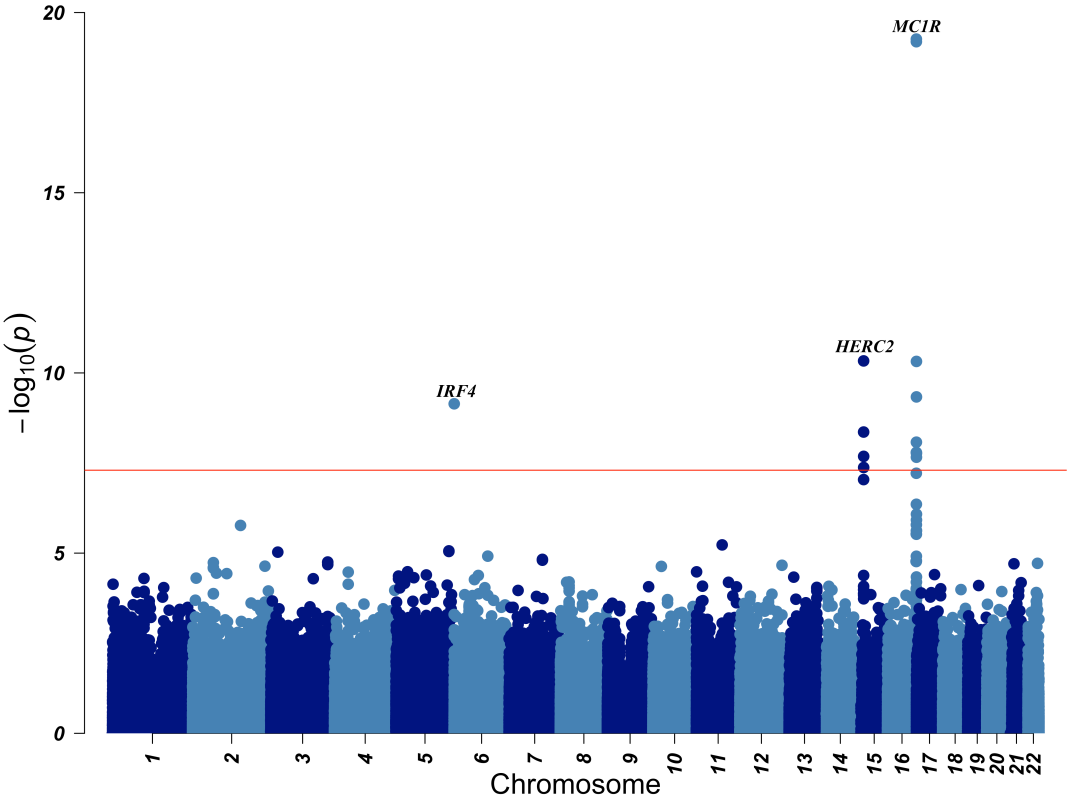
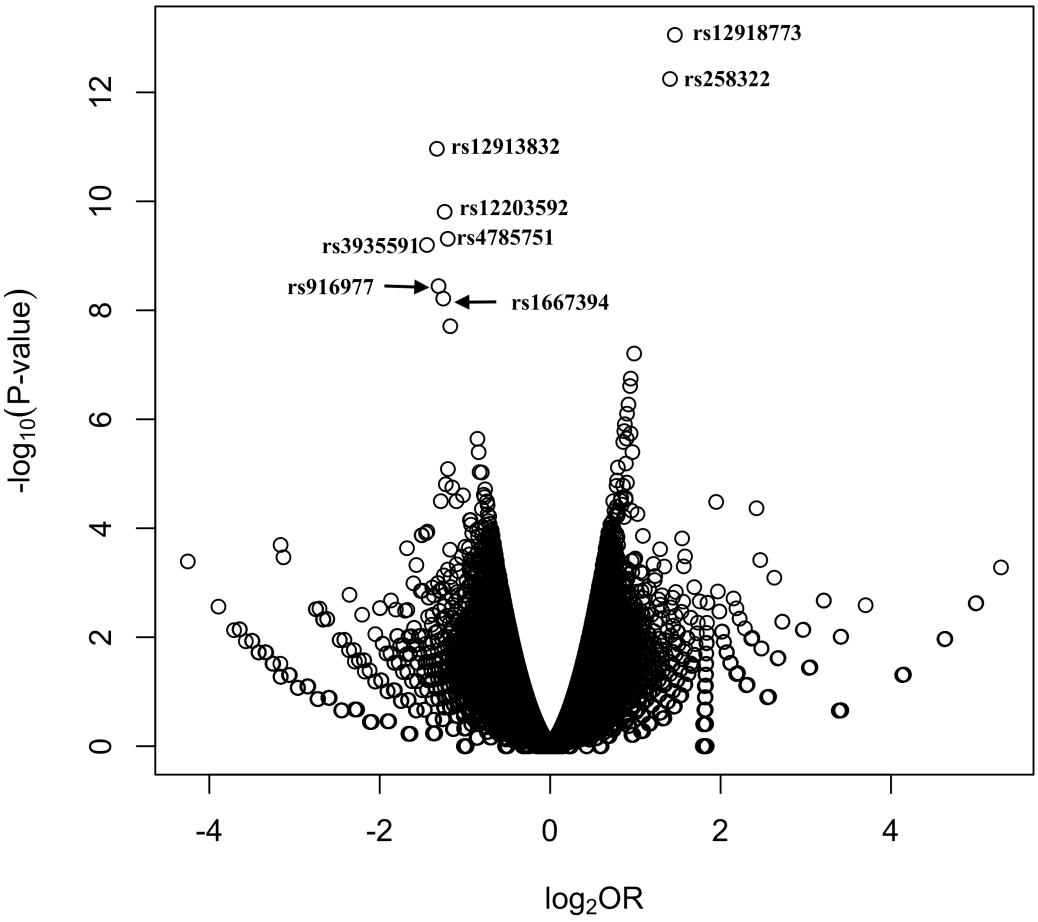


Figure 3.13 | Volcano plot for the GWAS of light versus dark hair. The observed – log_{10} transformed p-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).



3.3 Hair greying

Subjects that noticed hair greying early in their life (age < 35, n = 205, ~ 15% of the total samples) were tested against subjects that noticed late onset of hair greying (Age > 35, n = 1,210, Figure 3.3). The association analysis was performed using the Fisher's exact test (3 x 2 genotypic test) to test the genetic differences at 519,188 markers. A SNP on chromosome 5, rs7711077 ($P = 1.5 \times 10^{-7}$, $OR_{Rec} = 17.08$ minor allele positively associated with not greying early, Figure 3.14, Figure 3.15 and Appendix B.11) showed suggestive genome-wide significance ($P < 1 \times 10^{-5}$). rs7711077 lies in *NLN* which is a mitochondrial protein playing a key role in protein degradation, as part of protein quality-control (Quiros et al., 2015). Neurolysin cleaves neurotensin, a neuropeptide which modulates central dopaminergic and cholinergic circuits, and it may have a role in secretion of α-melanocyte-stimulating hormone (α-MSH) (Brown et al., 2001; Pan et al., 1992). The Q-Q plot is shown in Figure 3.16.

3.3.1 *IRF4* and hair greying

A study (Adhikari et al., 2016) has shown that the SNP rs12203592 is associated with increased hair greying. rs12203592 lies in an intron of interferon regulatory factor 4 (*IRF4*) which interacts with microphthalmia-associated transcription factor (MITF) to regulate tyrosinase (an enzyme in melanin synthesis) expression (Praetorius et al., 2013). rs12203592 showed genome-wide significance with black, blond and light versus dark hair colour in the current analysis of PoBI data ($P < 5 \times 10^{-8}$). rs12203592 showed suggestive significance (as *IRF4* is a candidate gene for hair greying) with hair greying in the current analysis ($P = 4 \times 10^{-5}$, $OR_{Rec} = 3.16$ minor allele positively

associated with not greying, Figure 3.14, green dot on chromosome 6, Table 3.11), replicating the previous result.

Table 3.11 | Distribution of rs12203592 genotypes in participants with hair greying phenotype.

Genotype	***AA***	***Aa***	***aa***	***Total***
Grey hair	26	69	110	205
No greying	53	394	762	1209
Total	79	463	872	1414

$P_{Dom} = 0.0128$, $OR_{Dom} = 1.47$; $P_{Rec} = 1.8 \times 10^{-5}$, $OR_{Rec} = 3.16$

Figure 3.14 | Manhattan plot for the GWAS of hair greying. The observed $-log_{10}$ transformed p-values (Fisher's exact test) were plotted (y axis) against their chromosomal positions (x axis). The horizontal red line indicates the threshold for genome-wide significance ($P = 5 \times 10^{-8}$) while blue line indicates the threshold for suggestive genome-wide significance ($P = 1 \times 10^{-5}$)

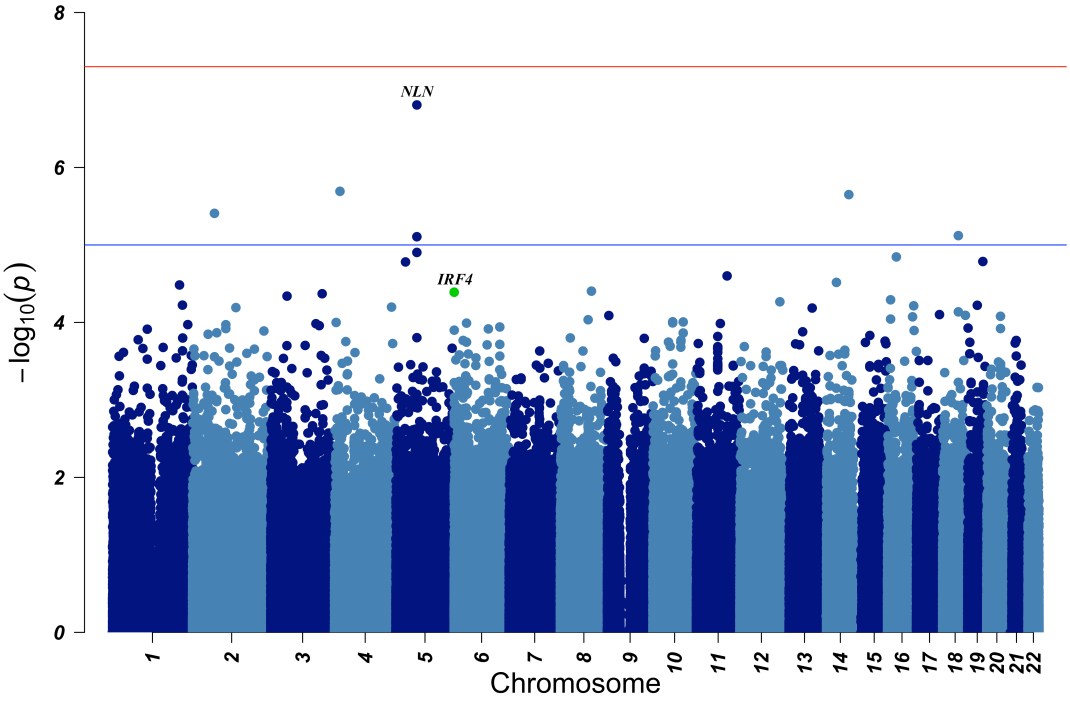


Figure 3.15 | Volcano plot for the GWAS of hair greying. The observed $-log_{10}$ transformed p-values (Fisher's exact test, recessive model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).

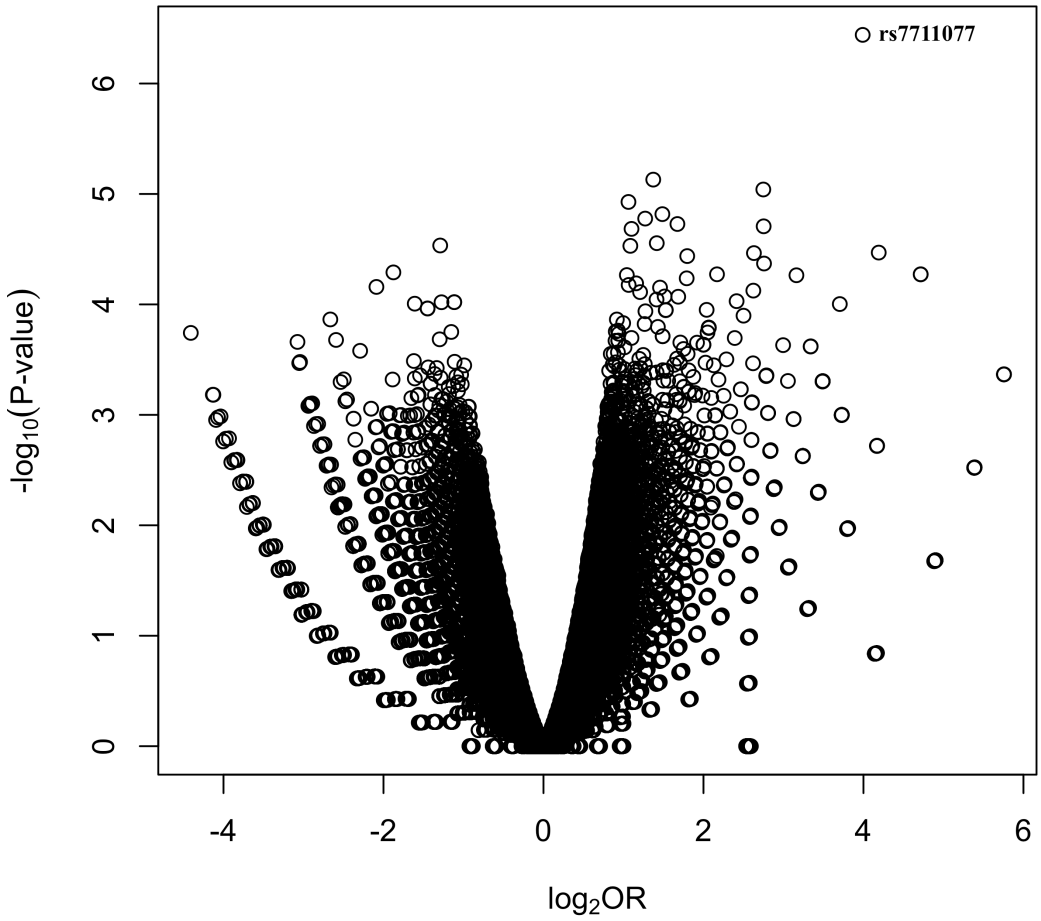
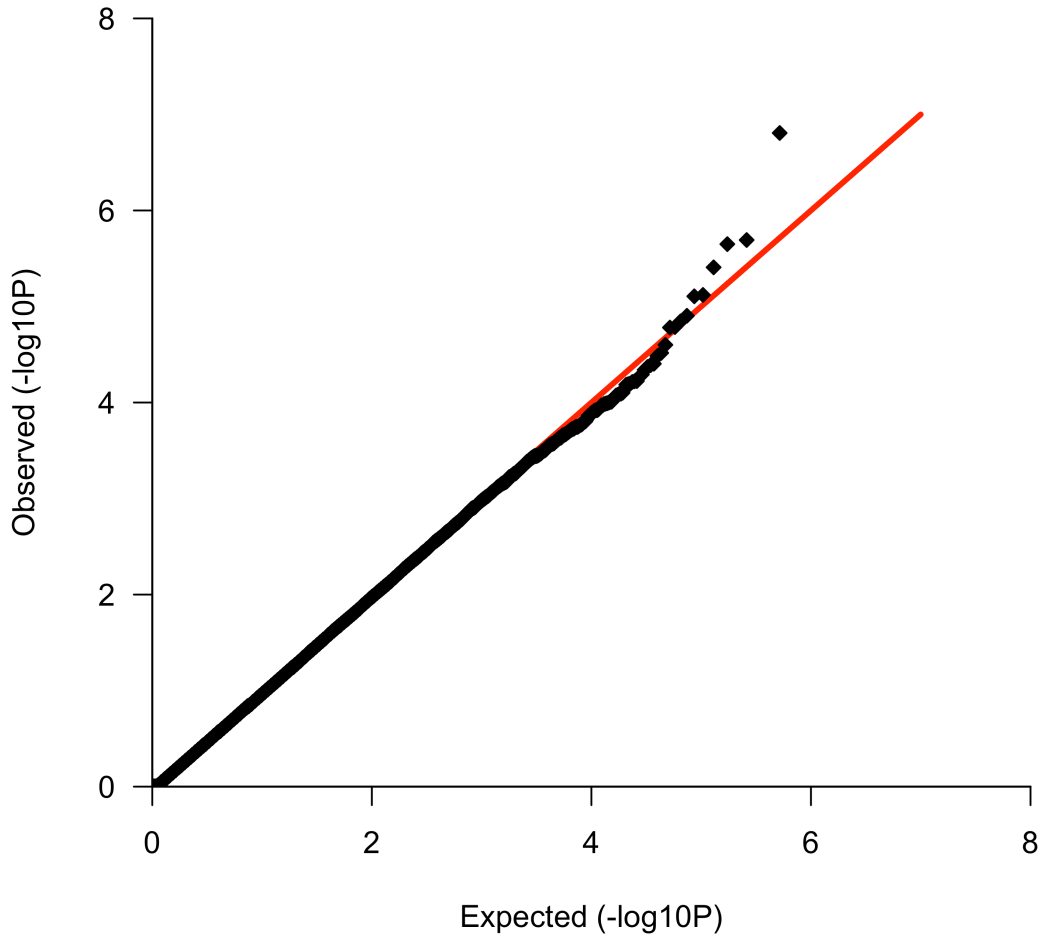


Figure 3.16 | Q-Q plot for the GWAS of hair greying. The $-log_{10}$ of observed p-values (Fisher's exact test) for each SNP were plotted against the $-log_{10}$ of expected p-values. The red line indicates the null hypothesis of no true association.



A brief summary of the most significant associations with hair pigmentation phenotypes along with characteristics of the data used for analysis is shown in Table 3.12. To show that the current data are consistent with previous studies and replicate the known associations for pigmentation, p-values and odds ratios for previously associated genetic variants are shown in Table 3.13.

Table 3.12 | A summary of genome-wide association signals of hair pigmentation observed in the PoBI data.

Hair Colour	***Cases***	***Controls***	***Locus***	***SNP***	***Gene[a]***	***P-value[b]***	***OR***
Red	91	1494	16q24.3	rs12918773	*SPATA33/MC1R*	1.0×10^{-26}	8.17
Blond	254	1098	6p25.3	rs12203592	*IRF4*	1.0×10^{-11}	3.15
Black	113	1472	15q13.1	rs12913832	*HERC2/OCA2*	6.7×10^{-11}	3.61
Light versus Dark	345	1211	16q24.3	rs12918773	*SPAT33/MC1R*	5.4×10^{-20}	2.75

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]Information from ensembl; [b]Fisher's exact test; OR, Odds ratio

Table 3.13 | Association statistics for known pigmentation variants observed in the PoBI data.

Locus[a]	***Chr Location***	***Variant***	***P-value***[b] ***(Red), (OR)***	***P-value***[b] ***(Blond), (OR)***	***P-value***[b] ***(Black), (OR)***	***P-value***[b] ***(Light Vs. Dark), (OR)***
MC1R	16q24.3	rs12918773	1.0×10^{-26}, (8.17)	6.6×10^{-5}, (13.21)	0.004052, (2.58)	5.4×10^{-20}, (2.75)
HERC2	15q13.1	rs12913832	0.8944	1.1×10^{-11}, (3.20)	6.7×10^{-11}, (3.61)	4.6×10^{-11}, (2.52)
IRF4	6p25.3	rs12203592	0.7092	1.0×10^{-11}, (2.90)	3.3×10^{-10}, (3.67)	7.1×10^{-10}, (2.36)
TYRP1	9p23	rs1408799	0.8655	0.5742	0.2228	0.4512
KITLG	12q21.32	rs11610915	0.4035	9.7×10^{-5}, (2.13)	0.07439	0.003016, (1.75)
TYR	11q14.3	rs1393350	0.3201	0.0385, (1.72)	0.9183	0.01944, (1.69)
SLC45A2	5p13.2	rs35391	1	0.6818	4.5×10^{-8}, (6.43)	0.2055
SLC24A4	14q32.12	rs12896399	0.1571	0.004198, (1.53)	0.2957	0.02407, (1.47)

Chr, Chromosome; [a]gene symbol from ensembl; [b]Fisher's exact test; OR, Odds ratio, Odds ratios are only shown for variants with p-value < 0.05

3.4 **Skin pigmentation**

Skin pigmentation was measured as the Melanin (M-index) and Erythema-index (E-index) using the DSM II skin colormeter (Cortex technology) by taking six measurements at randomly chosen points of the upper inner arm for each participant (3 from the left arm and 3 from the right arm, Appendix A.2). The M-index measures darkness and defined as $\log_{10}$ (1/% red reflectance) while the E-index measures redness defined as $\log_{10}$ (1/% green reflectance) - $\log_{10}$ (1/% red reflectance) (Candille et al., 2012; Shriver & Parra, 2000). The upper inner arm was chosen as site of sampling to avoid the effect of sun exposure on the measurement.

For each arm, the mean was taken over all three measurements. The scatter plots in Figure 3.17 show the strong correlation between the measurements from each arm (r = 0.81 for M-index and r = 0.72 for E-index), giving an indication of the stability of these measurements. Figure 3.18 shows the distribution of melanin and erythema measurements after averaging both arms (M-index, mean = 29.23, SD = 3.68, E-index mean = 9.12, SD = 1.94). A cutoff value of <27 for M-index and <7 for E-index were used to test the extremes versus controls based on the histogram (Figure 3.18). Pigmentation measurements were also collected for the forehead (to study the tanning ability), but these could not be analysed due to time constraints.

Figure 3.17 | Scatter plots showing the distribution of quantitatively measured skin M-index and E-index for both arms a) showing M-index before outlier removal, b) M-index after outlier removal, c) E-index before outlier removal and d) E-index after outlier removal. The correlation coefficient (*r*) values are given in the inset.

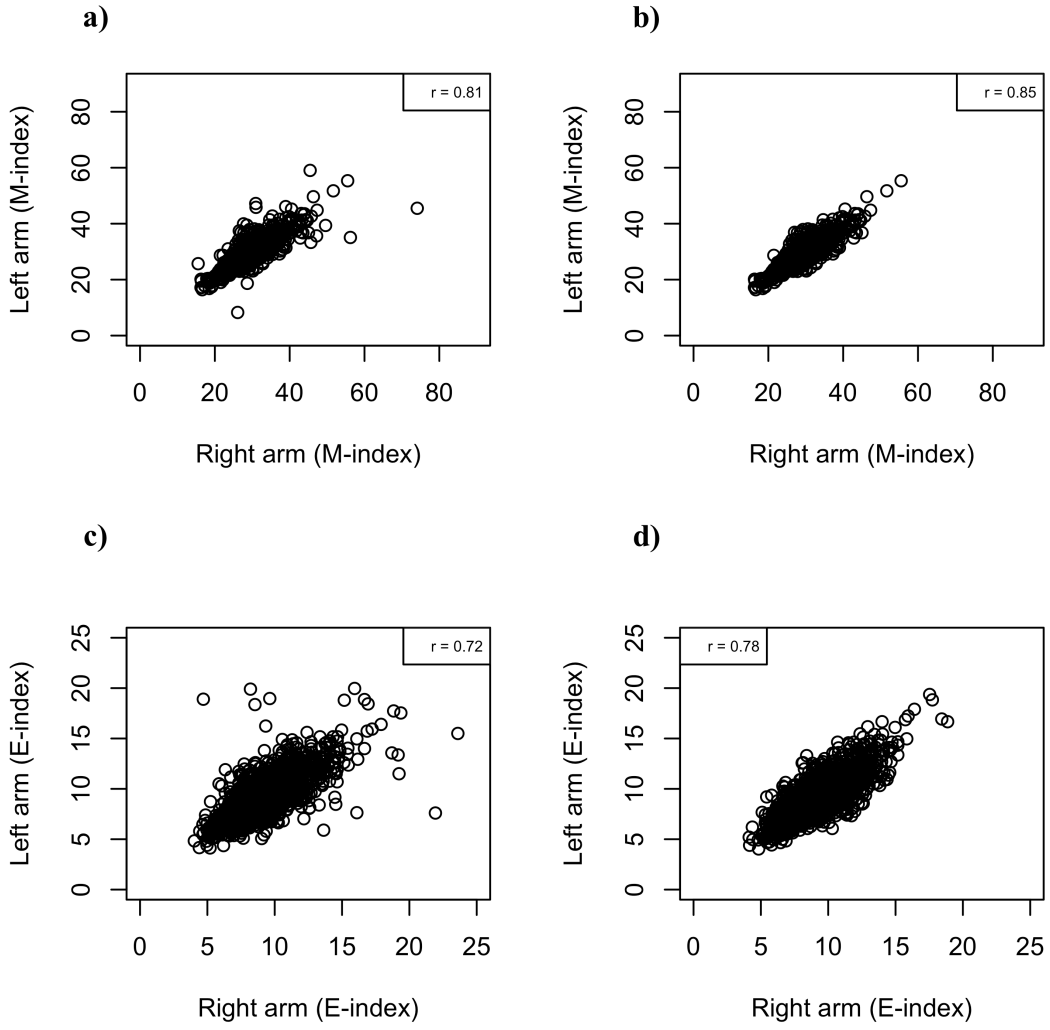
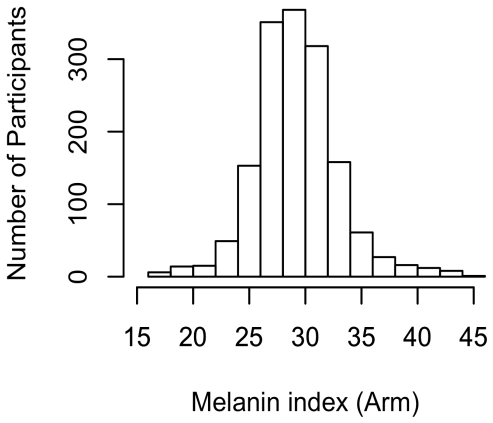
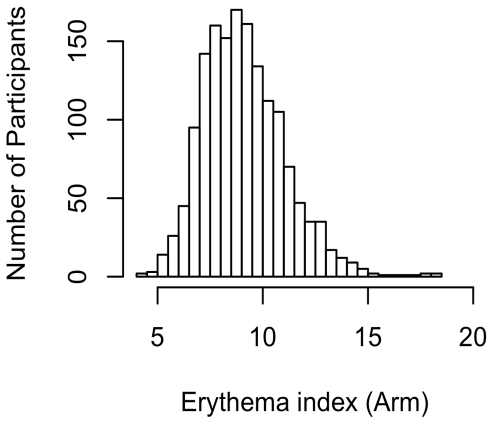


Figure 3.18 | Histograms showing the distribution of quantitatively measured a) Melanin-index, b) Erythema-index for both arms after excluding the outliers. Figure c shows the distribution of Melanin-index in four European population (Candille et al., 2012) measured with the same approach as in the PoBI data.

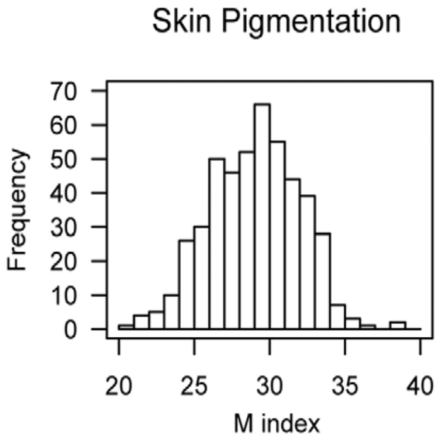
a)



b)



c)



3.4.1 Melanin-index

The associations between Melanin-index (586 subjects with M-index <27 and 971 subjects with M-index >27) and 519,188 markers were tested using the Fisher's exact test (3 x 2 genotypic test). All SNPs in the pigmentation candidate gene *SLC45A2* (Stokowski et al., 2007) showed statistical significance ($P < 0.05$, considered significant as *SLC45A2* is a candidate gene for skin pigmentation, Figure 3.19, green dots on chromosome 5, Table 3.14). In addition to SNPs within, rs4056123 ($P = 6.1 \times 10^{-7}$, $OR_{Dom} = 1.79$, Appendix B.12 and B.13), showed suggestive significance level. rs4056123, an intergenic variant, lies within 100 kb of *CHEK1* on chromosome 11. Checkpoint kinase 1 (*CHEK1*) is an important regulator of the cell-cycle and plays a fundamental role in DNA damage-induced cell-cycle arrest (Smith et al., 2013). Table 3.15 shows the novel genetic variants suggestively associated with skin pigmentation in the current study.

Table 3.14 | Association of *SLC45A2* variants with skin pigmentation in PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***p-value[b]***	***OR***	***r^2***
rs28777*	5	33958959	*SLC45A2*	0.0004198	2.78	-
rs35407	5	33946571	*SLC45A2*	0.001296	2.46	0.85
rs26722	5	33963870	*SLC45A2*	0.04783	2.11	0.44
rs35391	5	33955673	*SLC45A2*	0.02056	2.03	0.71
rs2278007	5	33951551	*SLC45A2*	0.02296	7.93	0.13

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with SNP rs28777); [b]Fisher's exact test; OR, Odds ratio; *see appendix B.14 for contingency table

A SNP on chromosome 12 (rs2384679, $P = 5 \times 10^{-6}$, $OR_{Rec} = 2.32$) showed suggestive significance with melanin level. rs2384679 lies in a long intergenic noncoding RNA (lincRNA).

Table 3.15 | Associations of novel genetic variants with skin pigmentation.

Phenotype	***SNP***	***Chr***	***Position***	***Gene[a]***	***P-value[b]***	***OR***
Melanin	rs4056123	11	125563373	-	6.1×10^{-7}	1.79
	rs2384679	12	116232428	-	5.0×10^{-6}	2.32
Erythema	rs3860254	12	80804476	*PTPRQ*	9.1×10^{-8}	2.26
	rs4341393	1	153776666	*GATAD2B*	3.1×10^{-7}	2.44
	rs1127091	1	153779412	*GATAD2B*	3.1×10^{-7}	2.44

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; [b]Fisher's exact test; OR, Odds ratio

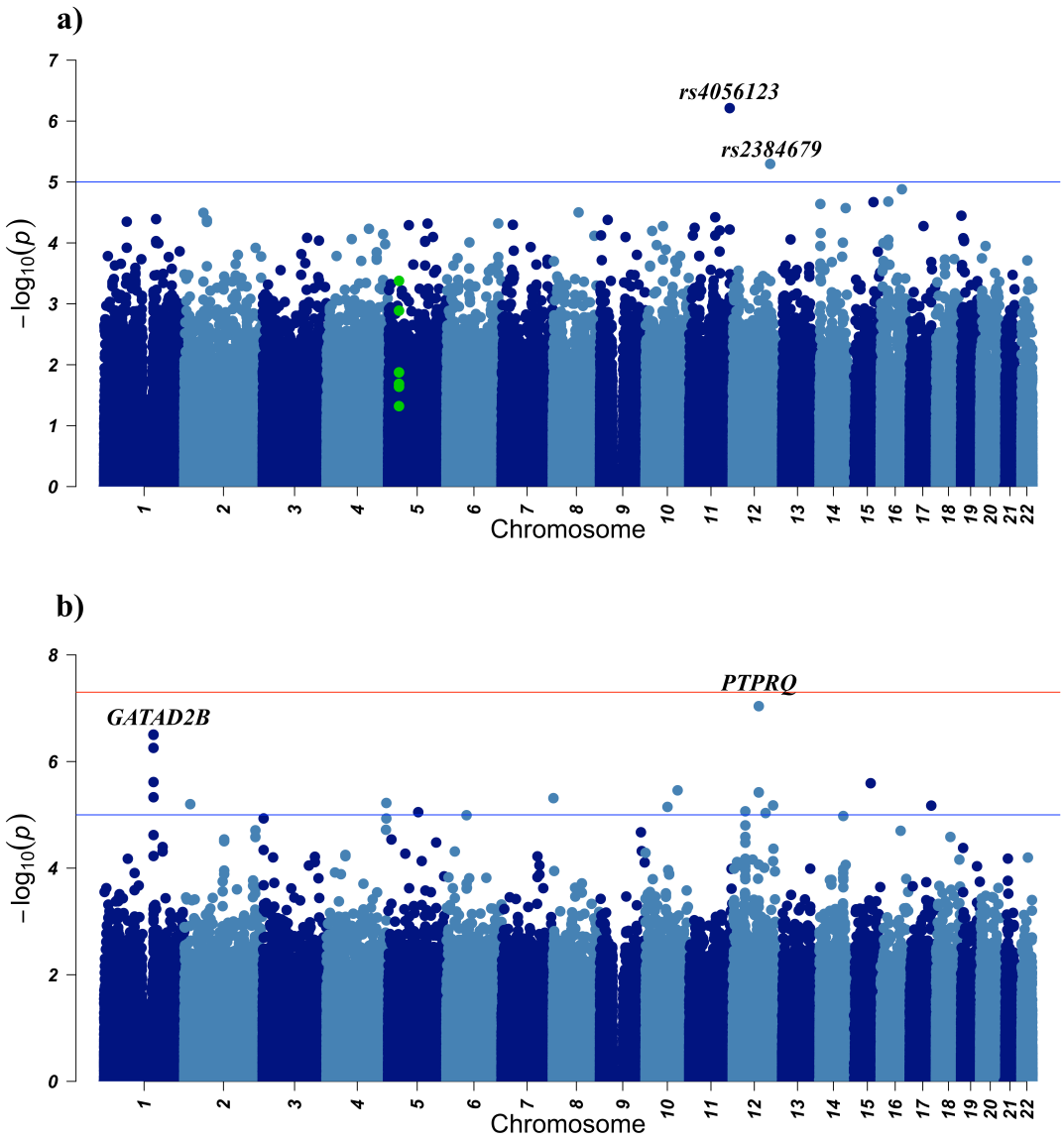
3.4.2 Erythema-index

For skin Erythema-index, rs3860254 ($P = 9.1 \times 10^{-8}$, $OR_{Rec} = 2.26$, Appendix B.12 and B.15) on chromosome 12 almost approached genome-wide significance ($P \leq 5 \times 10^{-8}$, Figure 3.19, Table 3.15). rs3860254 is an intron variant in *PTPRQ*. Protein tyrosine phosphatase, receptor type Q (*PTPRQ*) has been shown to have high phosphatidylinositol phosphatase (PIPase) activity compared to protein tyrosine phosphatase activity (Oganesian et al., 2003) and it is important to note that phosphatidylinositol phosphates such as PI(3,5)P2 have been shown to directly regulate the cation channel *TPC2* (two-pore-segment channel 2). This maintains both

melanosomal membrane potential, pH, and regulates pigmentation (Bellono et al., 2016).

Other than the SNP on chromosome 12, a cluster of SNPs almost approached genome-wide significance on chromosome 1. All SNPs (most significant SNP, rs4341393, $P = 3.1 \times 10^{-7}$, $OR_{Dom} = 2.44$, Appendix B.16) lie between the *SLC27A3* and *GATAD2B* region on chromosome 1. *SLC27A3 (*also known as fatty acid transporter 3) is known to mediate long-chain fatty acid transport and has been shown to express highly in epidermal keratinocytes (Schmuth et al., 2005). Another study has suggested that methylation patterns in *SLC27A3* may have a role in human cutaneous melanoma (Furuta et al., 2006).

Figure 3.19 | Manhattan plots of the quantitatively measured skin pigmentation GWAS. The green dots represent the SNPs in the candidate gene *SLC45A2*. The red and blue line indicates the genome-wide significance ($P = 5 \times 10^{-8}$) and suggestive genome-wide significance ($P = 1 \times 10^{-5}$), respectively.



3.5 **Discussion**

The present study reports genome-wide association analyses (GWAS) of hair pigmentation in a well-characterised population of the British Isles. Apart from confirming the previously known associated variants in *MC1R*, *HERC2*, *IRF4*, *SLC45A2* and *KITLG*, the current GWAS suggests novel variants (suggestively associated), which could play a role in pigmentation biology.

A SNP (rs9850626, $P = 2.6 \times 10^{-7}$, $OR_{Dom} = 4.90$) on chromosome 3 could potentially be a candidate SNP for future studies which aim to investigate more of these complex traits. rs9850626 in *KLF15* showed suggestive genome-wide significance with the red hair status in the current analysis. *KLF15*, a transcription factor, plays a key role in different cellular processes, including glucose metabolism, nitrogen homeostasis (Jeyaraj, Scheer, et al., 2012) and cardiac repolarization and arrhythmogenesis in mice (Jeyaraj, Haldar, et al., 2012). *KLF15* expression is controlled by *BMAL1* in cardiac repolarization. *BMAL1*, a circadian clock gene, has recently been shown to regulate melanin synthesis via the peripheral clock (Hardman et al., 2015). Importantly, rs9850626 did not reach suggestive significant for other hair colours tested in the current GWAS except red hair status. A literature survey found no study, prior to the current GWAS, indicating any role of this transcription factor in pigmentation biology. Expression and functional studies of *KLF15* in melanocytes would be critical before it can be concluded that *KLF15* has any significant role in melanin biology.

For the black versus non-black hair, genome-wide associations were observed on chromosomes 5, 6 and 15 confirming the known genetic variants that control pigmentation phenotypes. A SNP (rs2822994, $P = 1.1 \times 10^{-5}$ $OR_{Dom} = 2.45$) on chromosome 21 in *NRIP1* showed a novel association. Nuclear receptor interacting

protein 1 (NRIP1, also known as RIP140) has been shown to regulate the expression of circadian clock gene *BMAL1* (Poliandri et al., 2011), again a potential link indicating that circadian rhythm may have an important role in melanin synthesis.

Blond versus non-blond hair colour analysis replicated the known variants in *IRF4* and *HERC2* genes on chromosome 6 and 15, respectively. rs12821256 (P = 0.000128, OR_{Dom} = 1.76), which is shown to be associated with blond hair in northern Europeans (Guenther et al., 2014; Sulem et al., 2007), lies in *KITLG* and showed significant association with blond hair colour in the current PoBI data, although another SNP, rs11610915, in *KITLG* showed a much lower P-value ($P = 9.7 \times 10^{-5}$, considered significant given *KITLG* is a candidate gene for skin colour) and a higher odds ratio (OR_{Dom} = 2.13). A functional study has shown that rs12821256 underlies the molecular basis of blond hair (Guenther et al., 2014).

A SNP rs17643831 ($P = 2.1 \times 10^{-6}$, OR_{Rec} = 3.01) achieved suggestive genome-wide significance for association with blond hair status. This SNP, on chromosome 13 (q13.1), lies in *KL* (also known as *Klotho*), an ageing suppressor genetic factor, which has been shown to regulate MITF (microphthalmia transcription factor) and tyrosinase (TYR) expression (increased) in the primary cultured retinal pigment epithelium (Kokkinaki et al., 2013).

For the hair greying phenotype, rs7711077 ($P = 1.5 \times 10^{-7}$, OR_{Rec} = 17.08) almost reached genome-wide significance. rs7711077 is an intron variant in *NLN* whose product is a peptidase which cleaves neuropeptides such as neurotensin (Brown et al., 2001). Importantly, previous studies in mice have shown that neurotensin inhibits

alpha-MSH (a peptide hormone which controls pigmentation in humans) by regulating tuberoinfundibular dopaminergic neurons (Pan et al., 1992).

In addition, as discussed in the hair greying analysis, the SNP rs12203592 (*IRF4*) has been shown to be associated with hair greying in an admixed Latin American population (Adhikari et al., 2016), and replicated within the PoBI dataset.

In addition to the hair colour phenotypes, quantitatively measured skin pigmentation measurements were used for genome-wide association analysis in PoBI data. Although self-reported measurements or phenotypes are cheap and easy to collect, the error that is introduced due to subjectivity could result in loss of power to detect a genetic effect in GWAS (Sulem et al., 2007). For Melanin and Erythema-index, multiple measurements were taken and averaged to further reduce the error. In addition to replicating the well-documented pigmentation gene *SLC45A2* (Stokowski et al., 2007) ($P < 0.05$, Table 3.14), this study reports novel variants (suggestively associated), based on the large effect sizes and evidence from literature, that may have a role in the pigmentation biology. An intergenic variant (rs4056123, $P = 6.1 \times 10^{-7}$, $OR_{Dom} = 1.79$) within 100 kb of *CHEK1*, almost approached genome-wide significance for association with the melanin-index phenotype. *CHEK1* is an important regulator of the cell-cycle, which plays a key role in response to UV-induced DNA damage. A study has demonstrated that deletion of chk1 (an orthologue of human *CHEK1*) in mouse melanocytes results in a complete lack of coat colour and it was suggested that this protein is important for development of epidermal melanocytes (Smith et al., 2013).

For the erythema index phenotype, a SNP on chromosome 12 (rs3860254, $P = 9.1 \times 10^{-8}$, $OR_{Rec} = 2.26$) showed borderline genome-wide significance (Figure 3.19).

rs3860254 is an intron variant in *PTPRQ* which preferentially shows phosphatidylinositol phosphatase (PIPase) activity over protein tyrosine phosphatase (PTPase) activity (Oganesian et al., 2003). PTPRQ can dephosphorylate a number of phosphatidylinositol phosphates, which are important regulators of two-pore sodium channel in melanocytes (Bellono et al., 2016). Given that PTPRQ primarily dephosphorylates a number of phosphatidylinositol phosphates, it may have a role in maintaining the levels of phosphoinositides in the melanocyte membrane. Some members of the phosphoinositide family play an important role in pigmentation (Chow et al., 2007).

SNPs on chromosome 1 within *GATAD2B* and 100 kb of *SLC27A3* showed an association with the erythema phenotype. *GATAD2B* is a zinc-finger-containing transcription repressor and has not been shown to have any role in skin pigmentation. *SLC27A3* is a member of the solute carrier family 27 and encodes for a protein called fatty acid transporter 3 (FATP3). Expression studies in mouse have shown that FATP3 is heavily expressed by epidermal keratinocytes (Schmuth et al., 2005). It is important to note that although most fatty acids can be synthesized by keratinocytes, they lack the ability to synthesize essential fatty acids such as linoleic acid and alpha-linoleic acid (Khnykin et al., 2011). Both linoleic acid and alpha-linoleic acid have been shown to have pigment-lightening effects by suppressing melanin synthesis in melanocytes (Ando et al., 1998), and are popular as skin lightening products. Therefore, this evidence suggests that *SLC27A3* could play a major role in regulating the melanin synthesis mechanism by transporting the fatty acids into the cells.

3.6 **Conclusions**

The current GWASs of hair and skin pigmentation in a well-characterised population of the British Isles confirmed previously known genetic variants in *MC1R*, *SLC45A2*, *IRF4*, *HERC2* and *KITLG,* that regulate the synthesis of melanin in different biochemical reactions. This work also provides evidence (variants showing suggestive genome-wide significance with large effect sizes and a potential role based on the literature search) for novel genetic associations, including *KLF15*, *NRPI1, Klotho* (*KL*), *PTPRQ* and *CHEK1* which may have a major role to play in the genetics of melanin synthesis in humans.

4 Population genetics of taste perception

4.1 Introduction

The genetics underlying individual taste preference in humans has been an active area of research since A. L. Fox discovered that humans vary in their ability to perceive the taste of chemical compounds like phenylthiocarbamide (PTC) (Fox, 1932). Our fundamental understanding of taste and olfaction behaviour comes from research performed over the past eighty years, including the pioneering work by R. A. Fisher (working with J. Huxley and E. B. Ford) on anthropoid apes to show that the proportion of tasters and non-tasters among chimpanzees were similar to human populations (Fisher, 1939). It is now scientifically evident that humans can perceive several hundred chemically diverse compounds that feature differing proportions of bitter, salty, sour, sweet and umami (for example, taste of glutamate) tastes (Kinnamon & Margolskee, 1996; Reed et al., 2006). Variation in individuals' sensitivity to taste these compounds and the ability to distinguish between them is known to be genetic (Reed et al., 2006) and this variation may affect their dietary preferences (Smith et al., 2016); affecting overall health in turn e.g. due to effects of nutrition on weight gain, as it is likely to reflect perception of the taste more generally (Tepper, 1999). Chemical compounds like phenylthiocarbamide (PTC), 6-*n*-propylthiouracil (PROP, a compound with a chemical structure similar to PTC) and sodium benzoate have been of key importance in our understating of human taste behaviour at both cellular and molecular levels. Taste sensitivity to PTC is a well-known Mendelian trait in population genetics (Wooding, 2006) and is correlated with

the ability to taste other bitter substances found in nature, such as cruciferous vegetables e.g. broccoli, Brussels sprouts (Keller et al., 2002).

PTC is a non-toxic, synthetic compound which tastes bitter to most individuals, and its chemical structure resembles poisonous plant alkaloids found in nature. It is therefore hypothesized that bitter taste sensitivity is of selective advantage in humans as it helps avoid ingestion of toxic compounds (Callaway, 2012; Wooding et al., 2004). The observations from R.A. Fisher's PTC taste experiment performed on the chimpanzees were among the first pioneering empirical observations to support the relationship between genetics, natural selection and evolution (Fisher, 1939).

Our understanding of the molecular basis of mammalian taste, olfaction and food preferences has increased vastly since A. L. Fox and R. A. Fisher (with J. Huxley and E. B. Ford) made their observations (Fox, 1932). Studies involving functional expression of taste receptor genes and genome-wide association studies of taste sensitivity in humans have shaped our knowledge (Kim et al., 2003; Ledda et al., 2014). It is now clear that taste-receptor cells (TRC) in taste buds feature receptors that define individuals' taste sensitivities to diverse chemical compounds (Chandrashekar et al., 2006). Taste receptor, type 1 (T1R) and taste receptor, type 2 (T2R) are families of G protein-coupled receptors (GPCRs) which mediate sensitivity to sweet and bitter taste, respectively (Margolskee, 2002). G protein-coupled receptors (GPCRs) are the most diverse group of membrane proteins and play a key role in downstream signal transduction of different taste stimuli (Chaudhari et al., 1996; Kinnamon, 2012). A small family of GPCRs- T1R1, T1R2 and T1R3 mediate the sensitivity to sweet taste (Nelson et al., 2001) while bitter taste sensitivity is mediated by ~40 highly divergent GPCRs known as T2Rs (Chandrashekar et al., 2000). The gene that codes for bitter taste receptor and sensitivity to PTC is on

chromosome 7 (q34), commonly known as *TAS2R38* (taste receptor, type 2, member 38) (Kim et al., 2003). Polymorphism in the *TAS2R38* gene defines an individual's status as a taster or non-taster. This locus on chromosome 7 (q34) is known to mediate bitter taste sensitivity, although it could be specific to PTC and PROP (6-*n*-propylthiouracil). Some genome-wide association studies using caffeine as a tasting compound have demonstrated that taste receptors on chromosome 12 are associated with perceived bitterness of caffeine (Ledda et al., 2014) and these findings add further complexity to the mechanisms of bitter taste genetics. Although PTC and PROP will be of fundamental interest in population genetics studies of taste perception, taste sensitivity to other bitter compounds like caffeine and quinine etc. would also be of a key significance in future research.

4.2 GWAS of bitter taste perception

4.2.1 Introduction

Phenylthiocarbamide (PTC) and 6-*n*-propylthiouracil (PROP) tasting is one of the most studied Mendelian traits in human populations. Since A. L. Fox first discovered the variation in PTC sensitivity, studies have been performed to validate and characterize the molecular basis of T2Rs as the taste receptors. The functional expression of T2Rs and the role of G protein-coupled receptors in downstream signalling of bitter taste perception have been already identified (Chandrashekar et al., 2000). The present study performs genome-wide association analysis of bitter taste perception within a well-characterised population of the British Isles to validate what is already known, and to scan the genome further to see if Britain differs from populations previously studied for PTC and PROP tasting.

4.2.2 Phenotype data collection

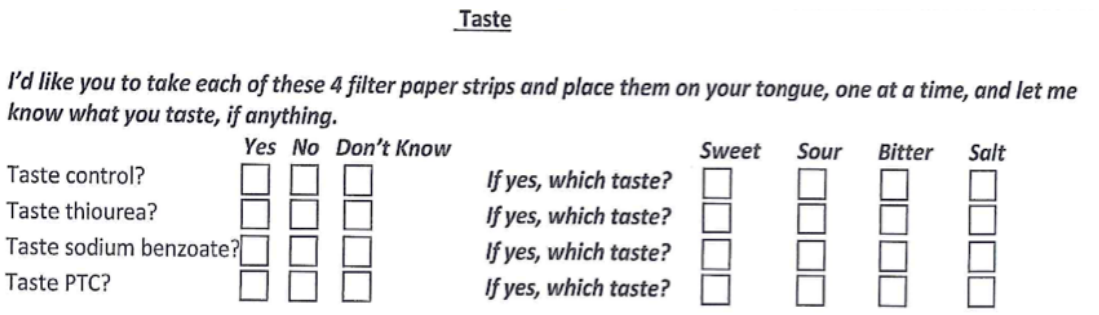
4.2.2.1 Subjects

Samples were collected from throughout the British Isles with a particular focus on samples with all four grandparents coming from the same area in rural Britain (Winney et al., 2012). Phenotype data (taste perception) from 2,136 samples were collected, of which 1,593 had genotype data available (n = 543 with no genotype data yet available). Participants were aged between 20-96 years (mean age 65.57 years). Each participant completed a written consent form and was led through a questionnaire about their perception of different tastes.

4.2.2.2 Phenotype definition

Each participant was provided four paper strips containing the control, PROP, sodium benzoate and PTC (made by Precision Laboratories, USA), respectively, to test the taste sensitivity. Control paper strips, which were tasteless and identical in colour to PROP; sodium benzoate and PTC, were used to exclude individuals with general taste problems as quality-control before further analysis. Participants were instructed to place the papers on their tongue, one at a time, and their response was recorded on the form (Figure 4.1, thiourea is an alternative name for PROP). Participants were told that there are no correct answers and that tasting ability differs according to their genetic makeup. Based on the ability to taste the paper strips, participants were asked to distinguish between sweet, sour, bitter and salty tastes.

Figure 4.1 | A part of the form used for taste phenotype data collection.



Taste

I'd like you to take each of these 4 filter paper strips and place them on your tongue, one at a time, and let me know what you taste, if anything.

	Yes	No	Don't Know		Sweet	Sour	Bitter	Salt
Taste control?	☐	☐	☐	If yes, which taste?	☐	☐	☐	☐
Taste thiourea?	☐	☐	☐	If yes, which taste?	☐	☐	☐	☐
Taste sodium benzoate?	☐	☐	☐	If yes, which taste?	☐	☐	☐	☐
Taste PTC?	☐	☐	☐	If yes, which taste?	☐	☐	☐	☐

thiourea; 6-n-propylthoiuracil (PROP), PTC; phenylthiocarbamide

4.2.3 Quality Control

4.2.3.1 PTC

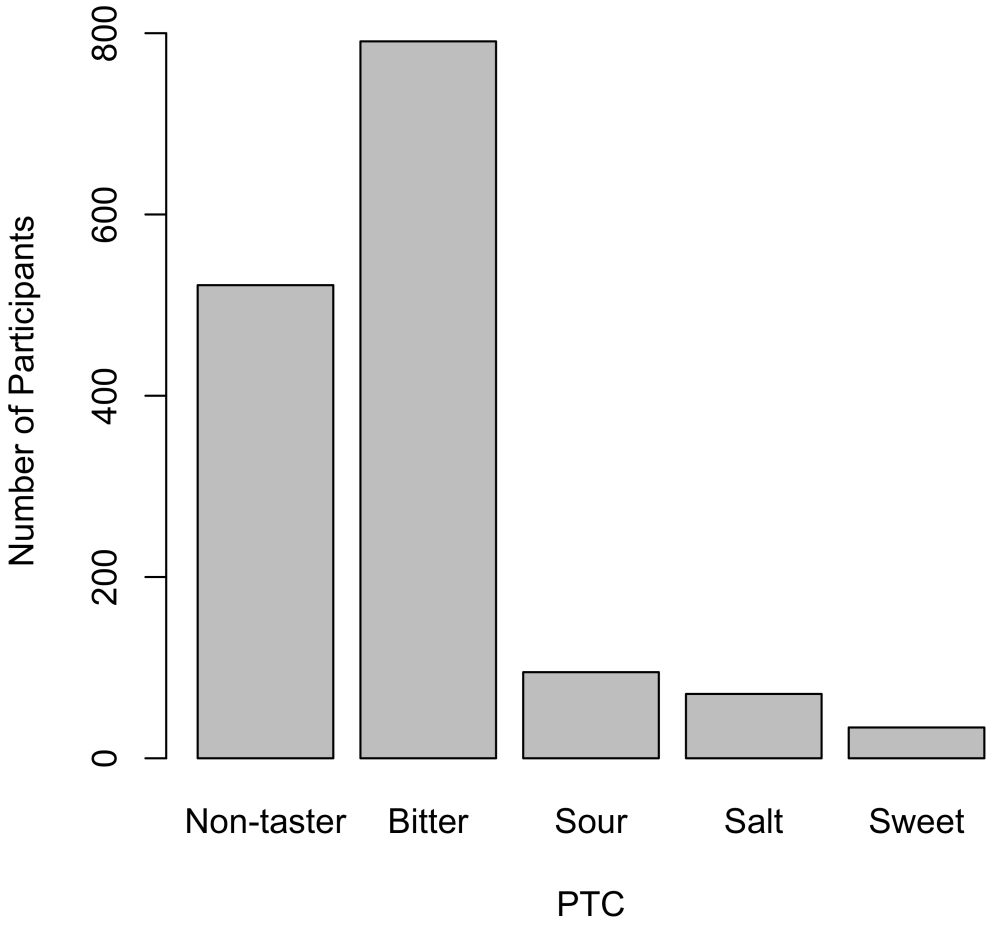
Before performing a genome-wide association analysis on the PoBI data for bitter taste perception, quality-control parameters were applied to the phenotype data. Out of 1,593 samples, which had genotype and phenotype data available, 1,513 participants (Figure 4.2) were eligible after excluding participants who tasted control paper strips (n = 80). Participants who responded as 'Don't know' were considered non-tasters. As PTC tastes bitter or tasteless (Kim & Drayna, 2005), participants who tasted PTC as sweet (n = 34) and salty (n = 71) were also excluded before further analysis (including the subjects with sweet and salty taste perception of PTC/PROP did not change the GWAS results except that the p-values for the most significant hits were higher, for the top hit - rs1726866, the p-value was $P = 1.9 \times 10^{-132}$ for PTC, and 3.4×10^{-32} for PROP. The same were applied to the variants which showed suggestive association with these traits). Final analysis was performed on 1,408 participants, and 886 tasters were tested against 522 non-tasters. Figure 4.2 (n = 1,513) and Table 4.1 (n = 1,408) show the characteristics of the samples used in this analysis.

Table 4.1 | Characteristics of 1,408 participants from the GWAS of PTC taste perception.

	Males	**Females**	**Mean age (SD)**
Tasters	391	495	64.88 (12.18)
Non-tasters	264	258	66.78(10.89)

SD; standard deviation, PTC; phenylthiocarbamide

Figure 4.2 | Distribution of PTC taste phenotype in the PoBI data.



4.2.3.2 PROP

PROP (6-*n*-propylthiouracil) is a chemically similar compound to phenylthiocarbamide (PTC, Figure 1.8 in the introduction chapter). Similar to PTC, individuals usually perceive PROP as bitter or tasteless. Figure 4.3 shows the number of participants and their responses for PROP taste sensitivity. Out of 1,513 samples, 117 participants were excluded from further analysis for tasting PROP as sweet (n = 39) or salty (n = 78). The final GWAS was performed on 1,396 (Table 4.2) participants, and 677 tasters were tested against 719 non-tasters.

Figure 4.3 | Distribution of PROP taste phenotype in the PoBI data.

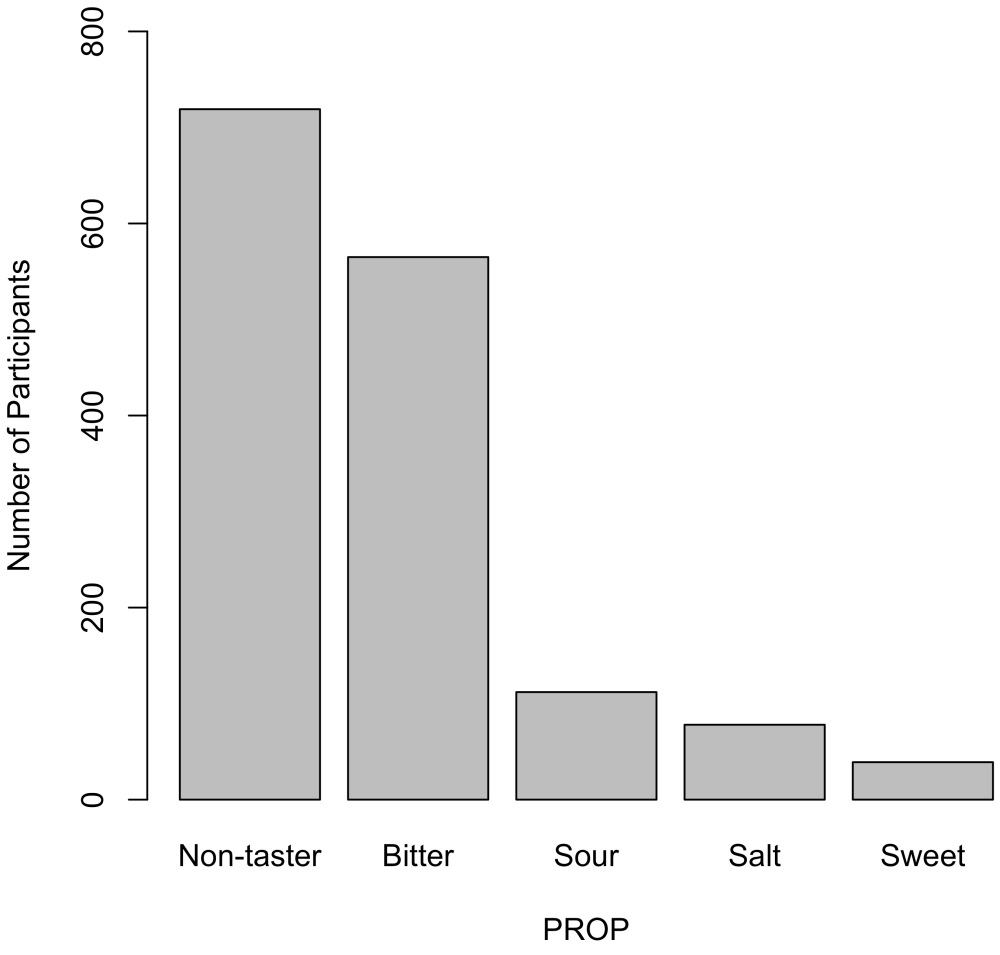


Table 4.2 | Characteristics of 1,396 participants from the GWAS of PROP taste perception.

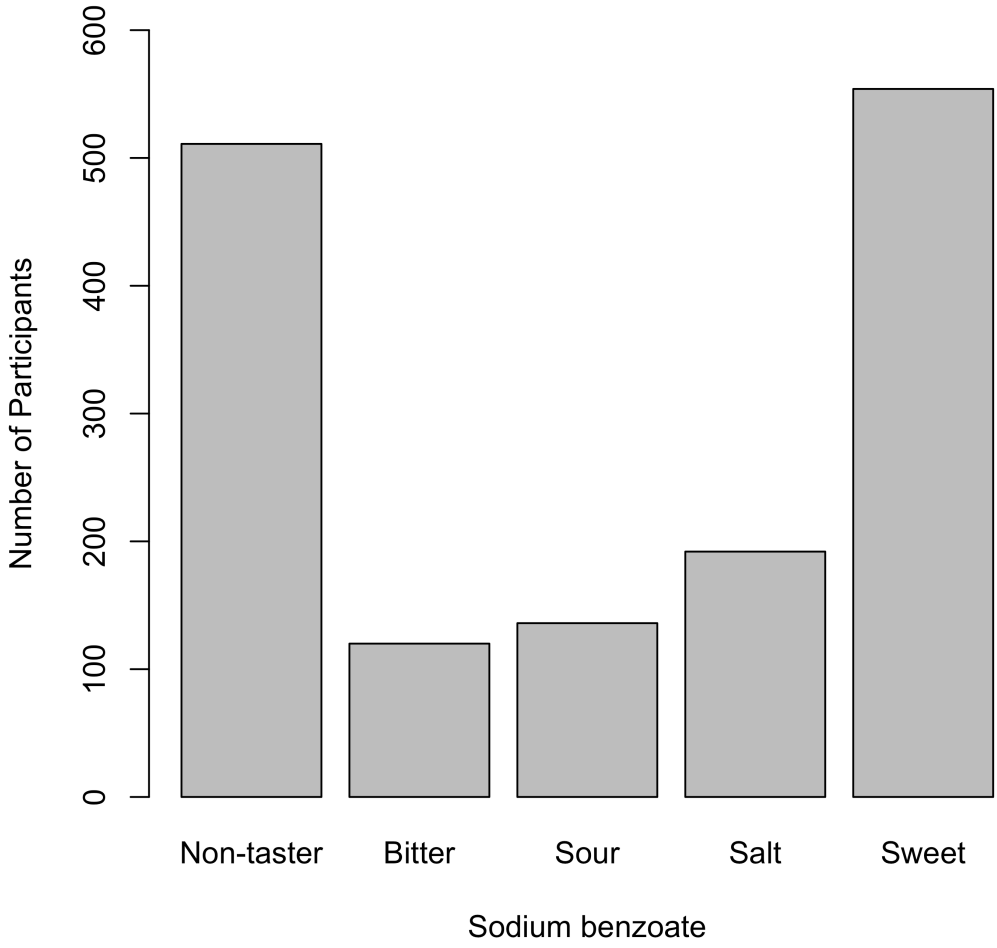
	Males	**Females**	**Mean age (SD)**
Tasters	268	409	64.54 (12.22)
Non-tasters	387	332	66.70 (11.24)

SD; standard deviation, PROP; 6-n-propylthoiuracil (PROP)

4.2.3.3 Sodium Benzoate

Sodium benzoate is commonly used as a food preservative. Unlike PTC and PROP, sodium benzoate doesn't produce one common taste and humans can vary highly between bitter, sour, salty, sweet or no taste in their response (Manley, 2015) (Figure 4.4). For sodium benzoate, participants who could taste strips in any way (tasters, n = 1,002, including sweet, sour, salt and bitter) were tested against non-tasters (n = 511). The mean age for tasters and non-tasters were 64.96 years (SD = 11.92) and 66.88 years (SD = 11.42), respectively.

Figure 4.4 | Distribution of sodium benzoate taste phenotype in the PoBI data.



4.2.4 GWAS of PTC bitter taste perception

The association between PTC taste sensitivity and genotypes at 519,188 single nucleotide polymorphisms (SNPs) was tested in 1,408 participants (tasters, n = 886, non-tasters, n = 522) using Fisher's exact test (3 x 2 genotypic test). Forty-two SNPs reached genome-wide significance ($P \leq 5 \times 10^{-8}$, Appendix C.1 for the complete list of SNPs). A clear peak on chromosome 7 characterises the Manhattan plot (Figure 4.5). Figure 4.6 shows the volcano plot of p-values and odds ratios of dominant

model. The most significant SNPs, rs1726866 ($P = 2.7 \times 10^{-140}$, $OR_{Dom} = 29.81$) and rs10246939 ($P = 6.7 \times 10^{-140}$, $OR_{Dom} = 29.73$, $r^2 = 1$) are among the three SNPs that form the main haplotypes behind the ability to perceive PTC as bitter (Campa et al., 2012; Kim et al., 2003). The third SNP of this haplotype (rs713598) was not genotyped in this dataset. The contingency table of genotypes for rs1726866 (along with information about the inheritance model), is shown in Table 4.3. An r^2 value of 1 (Table 4.4) between the SNP rs1726866 and rs10246939 reflects complete linkage disequilibrium (LD). The regional association plot (Figure 4.7) of chromosome 7 (q34) shows that the strongest association (rs1726866) lies within the *TAS2R38* gene confirming previous results (Kim et al., 2003).

Table 4.3 | Distribution of rs1726866 genotypes in PTC tasters and non-tasters.

Genotype	***AA***	***Aa***	***aa***	***Total***
Non-tasters	32	121	369	522
Tasters	228	591	66	885
Total	260	712	435	1407

$P_{Dom} = 1.4\times10^{-140}$, $OR_{Dom} = 29.81$; $P_{Rec} = 1.8\times10^{-22}$, $OR_{Rec} = 5.3$

Figure 4.5 | Manhattan plot for the GWAS of PTC taste perception. The observed – log_{10} transformed p-values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). The horizontal red line indicates the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

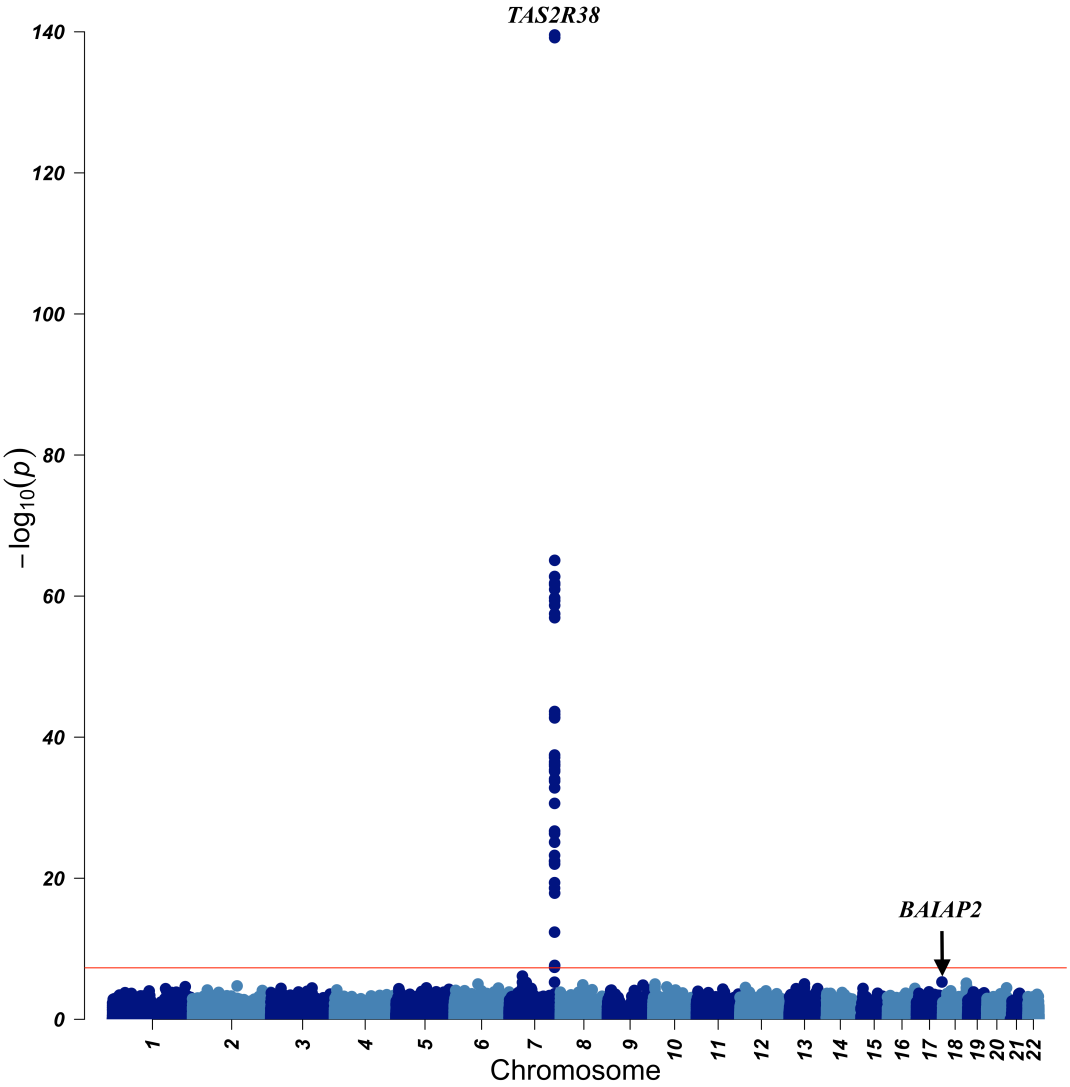


Figure 4.6 | Volcano plot for the GWAS of PTC taste perception. The observed – log_{10} transformed p-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).

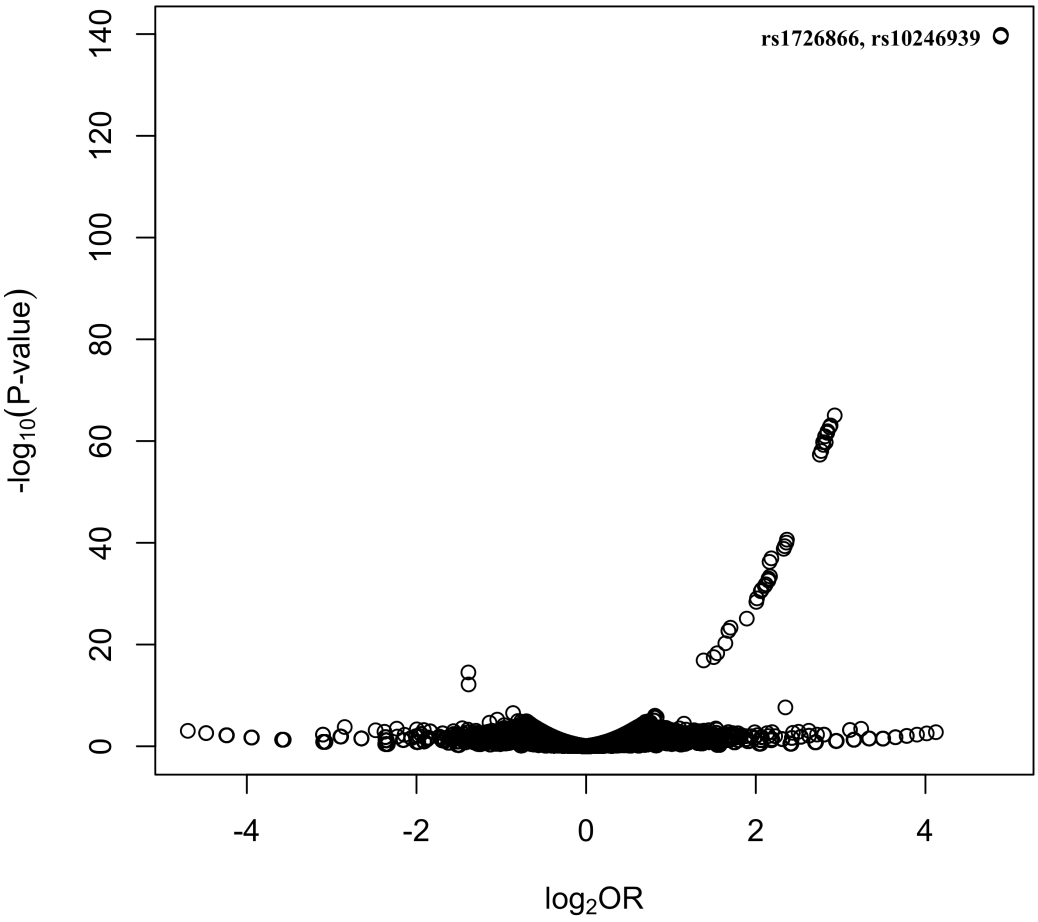


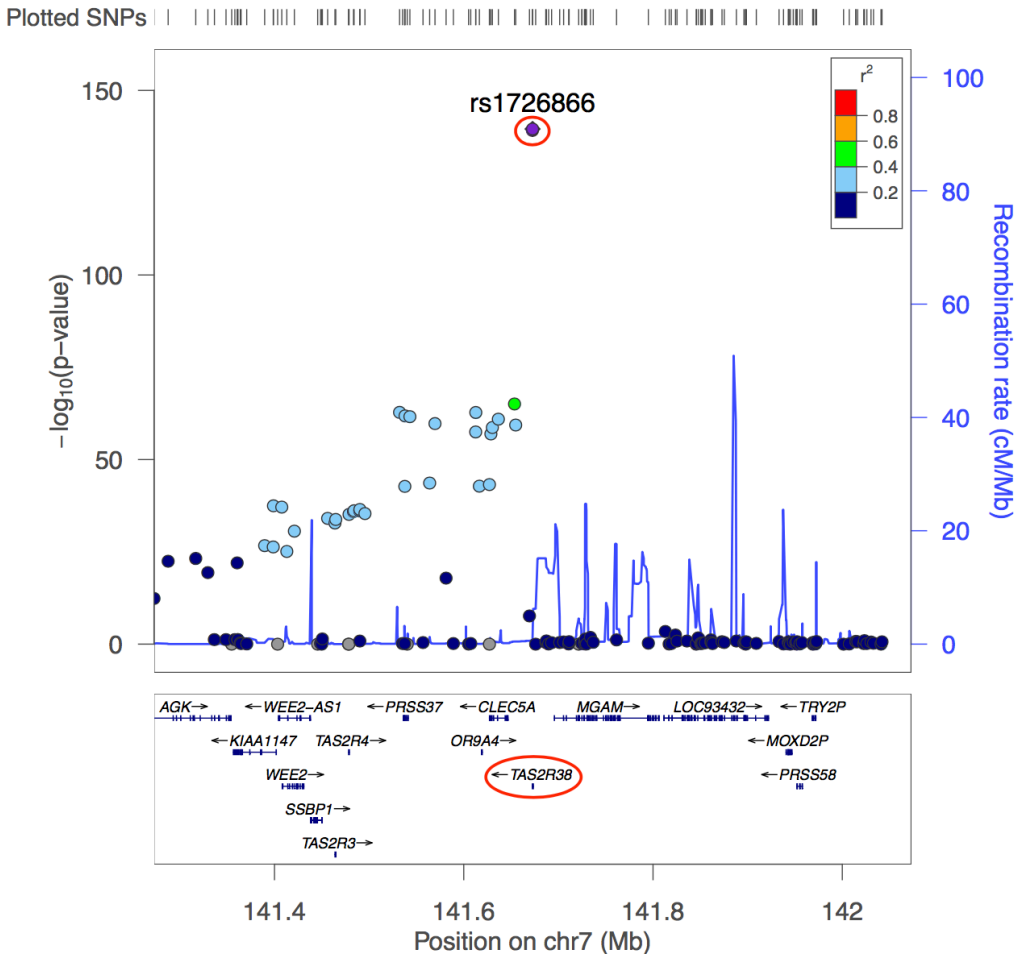
Table 4.4 | Association of genetic variants with PTC taste perception in the PoBI data.

SNP	***Chr***	***Position***	***gene[a]***	***P-value[b]***	***OR***	***r^2***
rs1726866	7	141672705	*TAS2R38*	2.7×10^{-140}	29.81	-
rs10246939	7	141672604	*TAS2R38*	6.6×10^{-140}	29.73	1
rs7794708	7	141653637	*MGAM*	8.4×10^{-66}	7.64	0.49
rs10464444	7	141532187	*PRSS37*	1.6×10^{-63}	7.38	.44
rs4969384	17	79065650	*BAIAP2*	5.1×10^{-6}	1.75	-

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with top SNP rs1726866); [b]Fisher's exact test; OR, Odds ratio

To further investigate the associations observed in the Manhattan plot (Figure 4.5), $-log_{10}$ of observed and expected GWAS p-values were plotted in a Q-Q plot (Appendix C.2). The clear deviation from the expected p-values is an indication of true association with the PTC bitter taste phenotype. To check that the deviation in Q-Q plot is because of statistically significant hits and not merely a systematic bias, e.g. due to population stratification, the Q-Q plot was also drawn after excluding the most significant associations (Appendix C.2), and this was clearly the case. Therefore, this GWAS replicates the previous findings that bitter taste sensitivity to PTC is driven by genetic polymorphisms in the *TAS2R38* gene (Kim et al., 2003). The confirmation of already known associations for PTC taste sensitivity also validates the PoBI dataset.

Figure 4.7 | Regional association plot of chromosome 7 (q34). Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 7 (q34). The purple diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD (r^2) with the most significant SNP. Genes in a particular region are shown at the bottom.



Other than the previously known associations on chromosome 7, a SNP on chromosome 17 (rs4969384, $P = 5.1 \times 10^{-6}$, $OR_{Dom} = 1.73$, Appendix C.3) showed suggestive genome-wide significance and could be a potential candidate for future studies. rs4969384 lies in *BAIAP2*, which is a key regulator in cell-cell

communication and also plays a major role in insulin homeostasis (Abbott et al., 1999; Yeh et al., 1996). It is important to note that taste receptors have been shown to be expressed in the gastrointestinal tract (Rozengurt, 2006) and studies in the past have shown that it could potentially affect insulin metabolism by mediating nutrient assimilation and endocrine responses (Dotson et al., 2008). In fact, a study has shown that a TAS2R haplotype on human chromosome 12 regulated glucose and insulin homeostasis (Dotson et al., 2008).

A possible mechanism by which the taste receptors mediate glucose and insulin metabolism is via affecting the release of cholecystokinin (CCK), secretin, Glucagon-like Peptide-1 (GLP-1), and Glucose-dependent Insulinotropic Polypeptide (GIP) through intracellular Ca^{2+} levels (Dotson et al., 2008; Rozengurt, 2006). Future studies will be required to check if *BAIAP2* and *TAS2R38* receptors have any role in regulating glucose and insulin homeostasis before concluding a potential link between metabolic disorders such as type 2 diabetes and *TAS2R38*.

4.2.5 GWAS of PROP bitter taste perception

GWAS performed on 1,396 participants (n = 719 non-tasters, n = 677 tasters) using Fisher's exact test (3 x 2 genotypic test) identified thirty-six SNPs that reached genome-wide significance ($P \leq 5 \times 10^{-8}$, Figure 4.8 and Appendix C.1); all of them were centred on chromosome 7 (q34) and were within *TAS2R38*. The strongest associations were with rs1726866 ($P = 4.3 \times 10^{-38}$, $OR_{Dom} = 4.60$, Table 4.5 & Table 4.6) and rs10246939 ($P = 6.5 \times 10^{-38}$, $OR_{Dom} = 4.59$, Table 4.5). The volcano plot of PROP GWAS and the Q-Q plot of observed p-values are shown in Figure 4.9 and Appendix C.2, respectively.

In addition to associations with *TAS2R38*, rs17327624 on chromosome 7 attained

suggestive genome-wide significance. rs17327624 ($P = 6.4 \times 10^{-6}$, $OR_{Rec} = 2.85$) lies in ATP binding cassette subfamily B member 1 (*ABCB1*) and a study has shown that T2R signalling activates *ABCB1* in the intestinal cells (Jeon et al., 2011). rs10210235 on chromosome 2 (p13.3) almost approached genome-wide significance ($P = 6.2 \times 10^{-7}$, $OR_{Dom} = 1.69$, Appendix C.4) and could be a potential candidate for further studies. rs10210235 lies in *ATP6V1B1*, a V-ATPase, is involved in intracellular signalling, olfaction and other cellular processes such as synaptic transmission (Merkulova et al., 2015; Paunescu et al., 2012). ATPase (e.g. ecto-ATPase and V-ATPase) is of particular interest in the cell biology of taste buds as these enzymes regulate ATP metabolism required for transmission of taste signals from taste bud to sensory nerve fibres (Bartel et al., 2006; Vandenbeuch et al., 2013).

Additionally, a biologically relevant SNP on chromosome 13 (rs11616269, $P = 4.7 \times 10^{-6}$, $OR_{Rec} = 5.24$, Appendix C.5) showed relatively high significance, lying near (< 100 kb) *TM9SF2*. *TM9SF2* is a member of the transmembrane 9 superfamily member 2, which has been shown to be expressed in the murine olfactory epithelium (Kanageswaran et al., 2015).

Table 4.5 | Association of genetic variants with PROP taste perception in the PoBI data. Only three most significant SNPs on chromosome 7 and two SNPs with $P < 1 \times 10^{-5}$ are shown here (Appendix C.1 for the complete list).

SNP	***Chr***	***Position***	***gene[a]***	***P-value[b]***	***OR***	***r^2***
rs1726866	7	141672705	*TAS2R38*	4.3×10^{-38}	4.60	-
rs10246939	7	141672604	*TAS2R38*	6.5×10^{-38}	4.59	1
rs10464444	7	141532187	*PRSS37*	5.5×10^{-17}	2.49	0.44
rs10210235	2	141543098	*ATP6V1B1*	6.2×10^{-7}	1.69	-
rs11616269	13	100106887	*TM9SF2*	4.7×10^{-6}	5.24	-

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (LD with top SNP rs1726866); [b]Fisher's exact test; OR, Odds ratio

Figure 4.8 | Manhattan plot for the GWAS of PROP taste perception. The observed – log_{10} transformed p-values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). The horizontal red line indicates the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

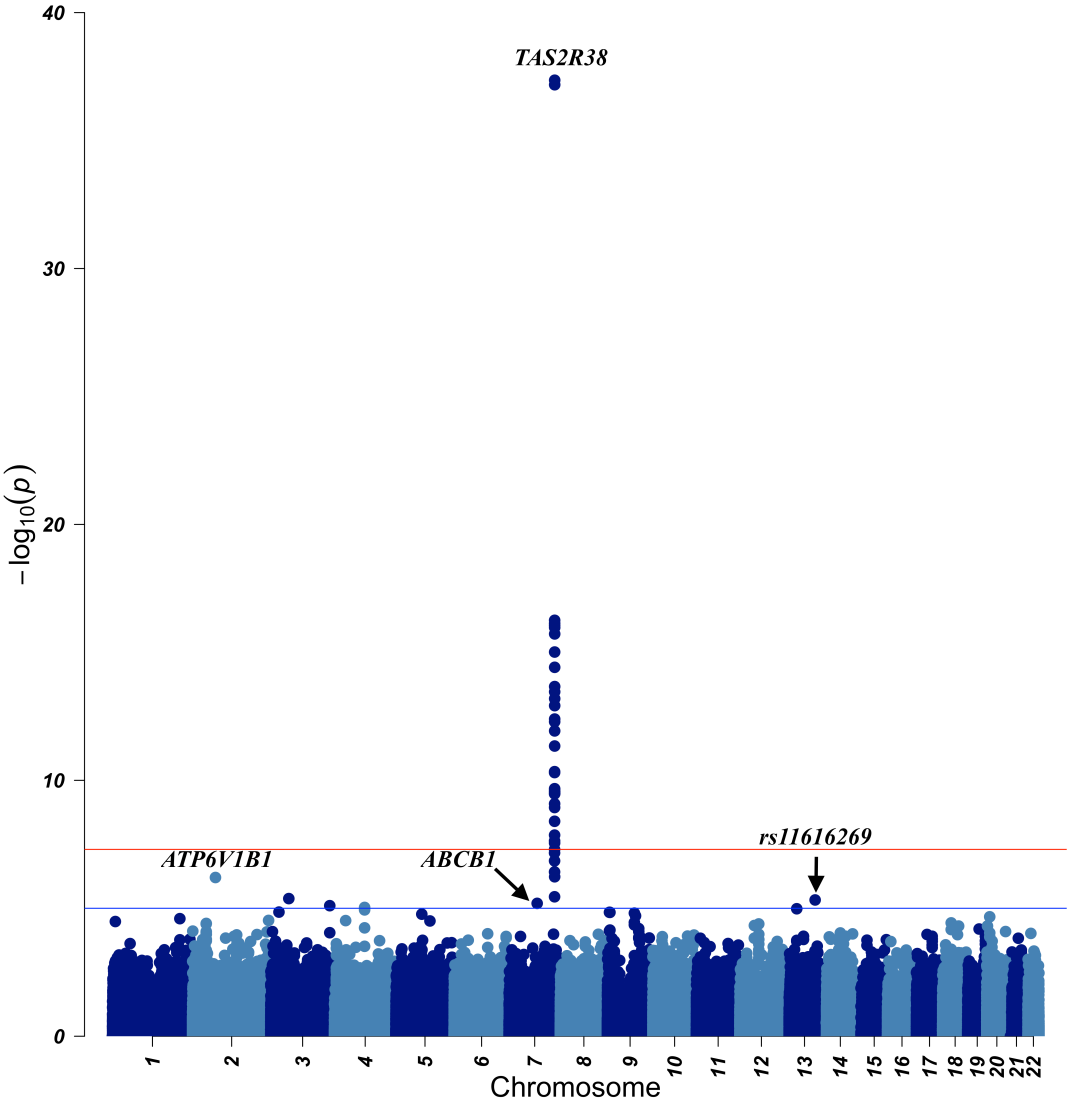
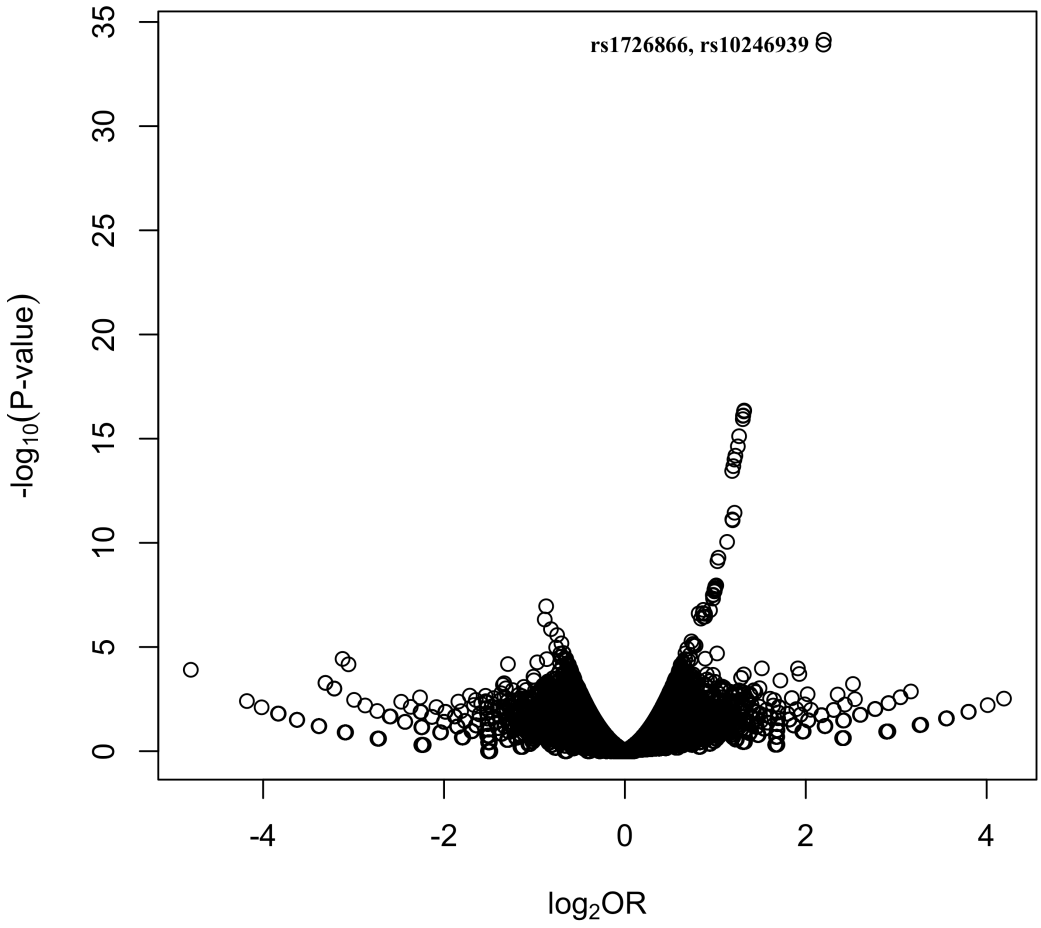


Table 4.6 | Distribution of rs1726866 genotypes in PROP tasters and non-tasters.

Genotype	***AA***	***Aa***	***aa***	***Total***
Non-tasters	77	320	321	718
tasters	185	391	101	677
Total	262	711	422	1395

$P_{Dom} = 7.2\times10^{-35}$, $OR_{Dom} = 4.60$; $P_{Rec} = 1.3\times10^{-15}$, $OR_{Rec} = 3.12$

Figure 4.9 | Volcano plot for the GWAS of PROP taste perception. The observed – log_{10} transformed p-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).



4.3 GWAS of sodium benzoate taste perception

For sodium benzoate taste perception, 1,002 tasters were tested against 511 non-tasters on 519,188 SNPs using fisher's exact test (3 x 2 genotypic test). As shown in the Manhattan plot (Figure 4.10), the strongest peak was observed on chromosome 1 (most significant SNP, rs3789421, $P = 6.7 \times 10^{-6}$, $OR_{Dom} = 1.44$) where four SNPs reached suggestive genome-wide significance ($P \leq 1 \times 10^{-5}$, Table 4.8). Table 4.7 shows the distribution of rs3789421 genotypes in tasters and non-tasters along with the odds ratios under dominant and recessive genetic models. rs3789421 implicates the ATP-binding cassette subfamily A member 4 (*ABCA4*) gene which has no previous proven evidence of association with taste perception.

Importantly, rs2645075 ($P = 6.7 \times 10^{-6}$, $OR_{Rec} = 4.16$, Figure 4.11 & Figure 4.12, Appendix C.6) is < 1 Mb (LD, $r^2 = 0.00094$) away from the taste receptor gene *TAS1R3* and previous research has shown that polymorphisms in the *TAS1R3* region are associated with saccharin preference in mouse strains (Reed et al., 2004). Similarly, another SNP on chromosome 1 (rs12141954, $P = 8.2 \times 10^{-6}$, $OR_{Rec} = 1.94$, Appendix C.7) lies within 1.4 Mb (LD, $r^2 = 0.00020$) region of taste receptor *TAS1R1*, which is particularly interesting as *TAS1R1* and *TAS1R3* receptors function in combination in sweet and umami taste perception (Zhao et al., 2003), although weak or no LD between the top SNPs and taste receptor gene variants implicate that it is an independent association. The Q-Q plot is shown in Appendix C.2.

Table 4.7 | Distribution of rs3789421 genotypes in sodium benzoate tasters and non-tasters.

Genotype	***AA***	***Aa***	***aa***	***Total***
Non-tasters	18	108	384	510
Tasters	12	310	678	1000
Total	30	418	1062	1510

$P_{Dom} = 0.002873$, $OR_{Dom} = 1.44$; $P_{Rec} = 0.003151$, $OR_{Rec} = 3.0$

Table 4.8 | Association of genetic variants with sodium benzoate taste perception in the PoBI data.

SNP	***Chr***	***Position***	***gene[a]***	***P-value[b]***	***OR***	***r^2***
rs3789421	1	94565577	*ABCA4*	6.7×10^{-6}	1.44	*
rs2645075	1	2254910	*MORN1*	6.7×10^{-6}	4.16	*
rs10158007	1	14575256	*KAZN*	8.0×10^{-6}	2.27	*
rs12141954	1	5342295	-	8.7×10^{-6}	1.94	*
rs10447360	6	71599826	*B3GAT2*	8.2×10^{-6}	1.65	-

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared, *r^2 values are not shown here as the SNPs are distant from each other; [b]Fisher's exact test; OR, Odds ratio

Figure 4.10 | Manhattan plot for the GWAS of sodium benzoate taste perception. The observed $-log_{10}$ transformed p-values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). The horizontal blue line indicates the threshold for suggestive genome-wide significance ($P = 1 \times 10^{-5}$).

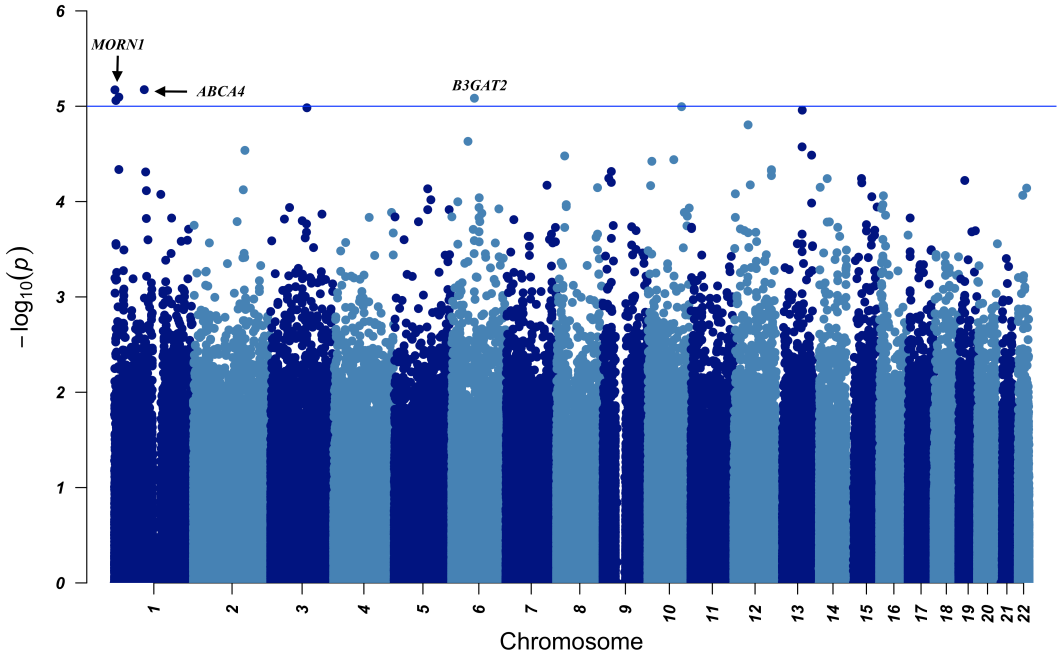


Figure 4.11 | Volcano plot for the GWAS of sodium benzoate taste perception. The observed $-log_{10}$ transformed p-values (Fisher's exact test, recessive model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).

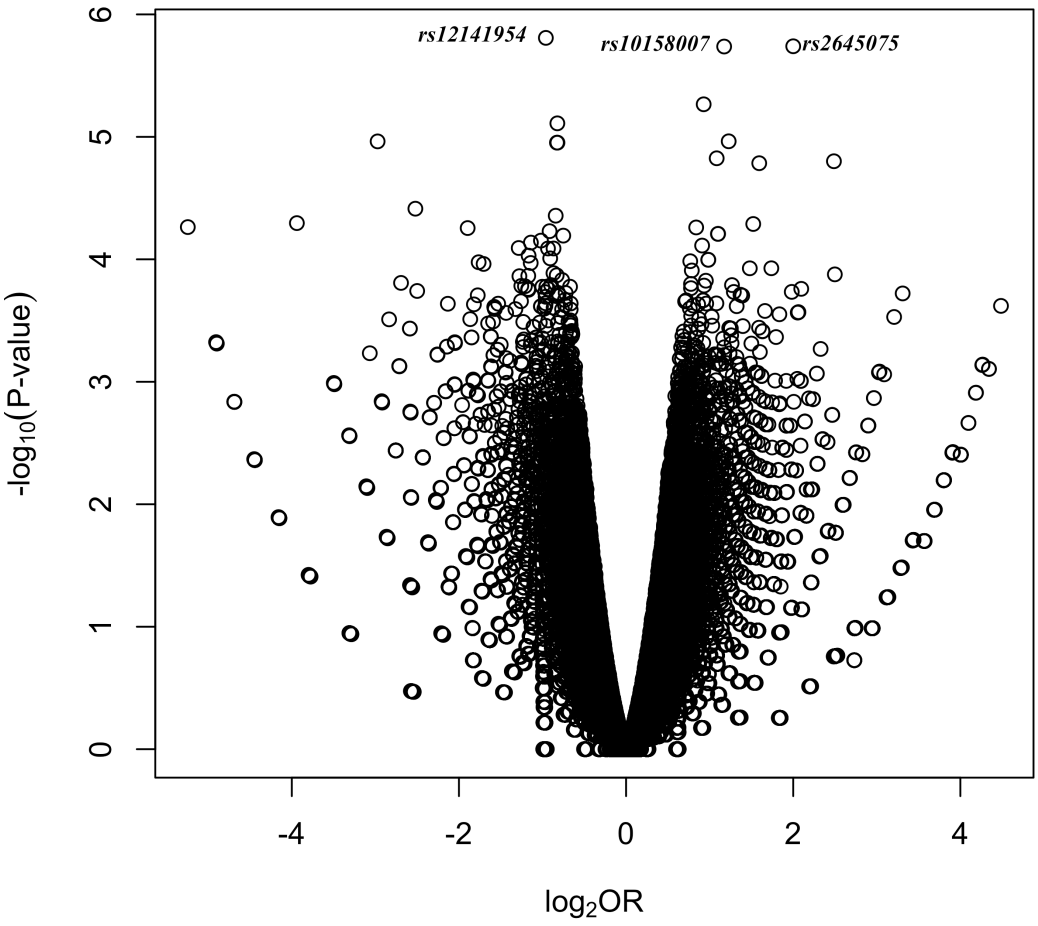
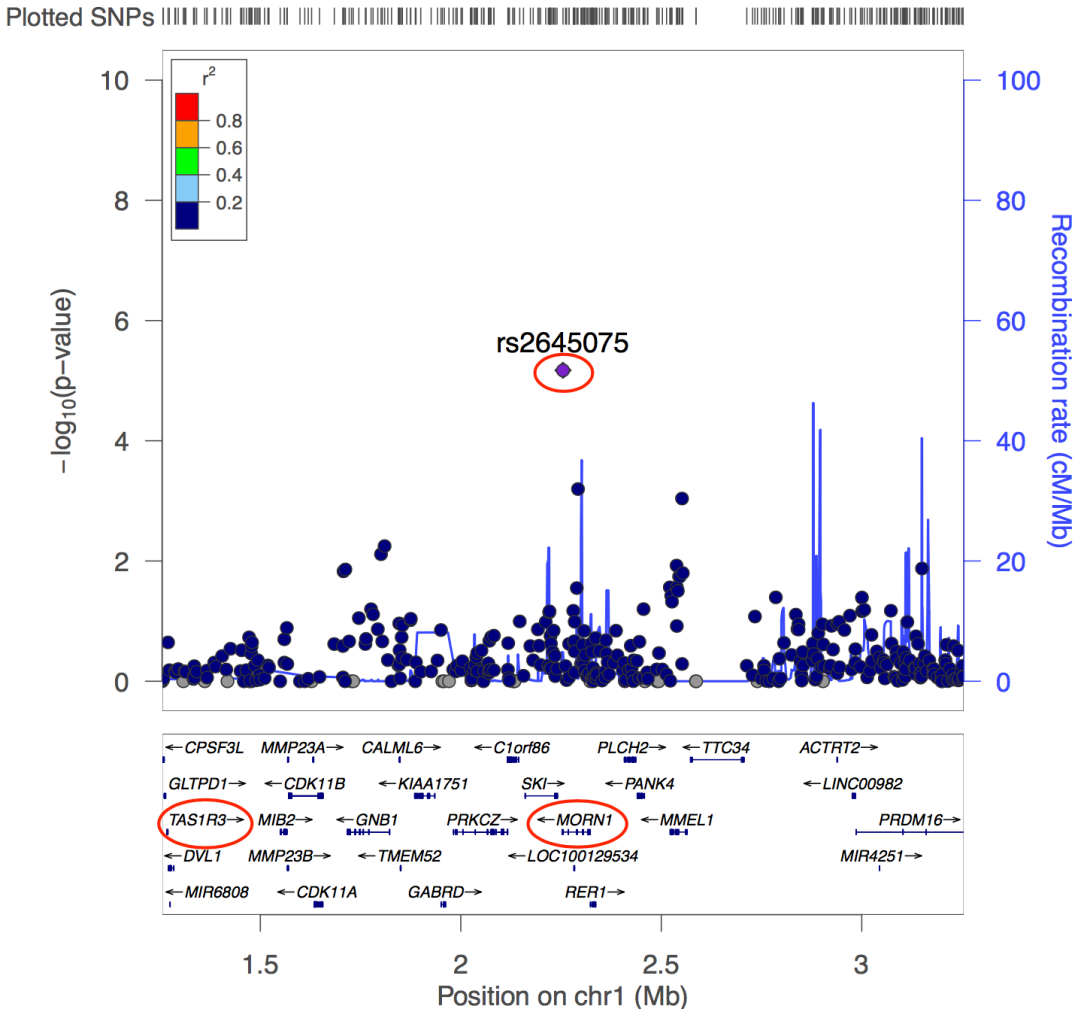


Figure 4.12 | Regional association plot of chromosome 1 (p36.33). Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 1 (p36.33). The purple diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD (r^2) with the most significant SNP. Genes in a particular region are shown at the bottom.



4.4 **Discussion**

Although polymorphism for PTC taste perception is well acknowledged and could be explained by variation in *TAS2R38* haplotypes, no previous research has been performed to analyse PTC taste perception in a well-characterised population like PoBI. The current study reports genome-wide significant hits for PTC; PROP and suggestive hits for sodium benzoate taste perception in a British population, confirming the known variant *TAS2R38* and suggesting novel variants which influence individual's ability to taste different compounds.

The SNPs rs1726866 and rs10246939 in the *TAS2R38* gene showed the strongest association with PTC and PROP bitter taste sensitivity and supported the previous findings (Kim et al., 2003). As shown by the r^2 values (Table 4.4 & Table 4.5), both rs1726866 and rs10246939 provide identical information, and are two of the three SNPs that form the bitter taste haplotype. A clear peak of SNPs was observed for both PTC and PROP as shown in the Manhattan plot (Figure 4.5 & Figure 4.8, respectively). For PTC, rs4969384 in *BAIAP2* on chromosome 17 (q25.3) also showed suggestive genome-wide significance. Although literature is not available relating *BAIAP2* with taste perception, it may have a potential role in taste biology, but further studies will be needed to draw any conclusions.

For the PROP taste perception, in addition to the *TAS2R38* variants, rs10210235 showed high, if not genome-wide significance. rs10210235 is an intron variant in *ATP6V1B1* (V-ATPase B1 subunit), and lies on chromosome 2 (p13.3). Studies have shown that loss of *ATP6V1B1* subunit results in impaired olfactory function in mouse (Paunescu et al., 2012). Therefore, *ATP6V1B1* could be a genetic component in taste

and olfaction, and further studies are needed to provide more information on its potential role in human sensory perception. It would be interesting to scan variants in the 1000 Genomes data with high LD in *ATP6V1B1* for functional candidates.

For sodium benzoate, no SNPs reached genome-wide significance level; however, the highest peak was on chromosome 1 (p22.1). The most significant SNP in this region was rs3789421, which lies in *ABCA4,* and has no-known evidence of active biological roles in the perception of taste. The absence of highly genome-wide significant associations for sodium benzoate taste perception, relative to the PTC and PROP, could be due to small effect sizes of different variants and the sample size of the current study thus lacking statistical power to detect them. It could also be possible that sodium benzoate taste ability is less heritable and Gene-Environment (GXE) or Gene-Gene (GXG) interactions play a major role in determining an individual's sensitivity to taste sodium benzoate.

It is known that the combination of *TAS1R1* and *TAS1R3* taste receptors form a G-protein-coupled receptor complex that mediates sweet and umami taste sensitivity (Nelson et al., 2001). As discussed earlier, rs12141954 (near *TAS1R1*) and rs2645075 (near *TAS1R3*) showed suggestive genome-wide significance (OR range 1.94 - 4.16) for association with sodium benzoate tasting (although they were not in LD). Notably, sweet is the most commonly reported taste for sodium benzoate in the PoBI dataset. Taken together, current research confirms PTC and PROP bitter taste sensitivity variants and suggest potential markers which could be of particular interest in human sensory perception.

4.5 **Conclusions**

Although a lot of research has been performed to understand the genetic polymorphisms behind taste perception, no study prior to this was reported in a well-characterised population of the British Isles. Apart from confirming the known genetic variants for taste perception, this GWAS also suggests *BAIAP2* and *ATP6V1B1* as novel candidate genes (based on suggestive genome-wide significant p-values and relatively large odds ratios) for future studies in human taste genetics.

5 Genetics of food preferences

5.1 Introduction

There has been a growing interest in the field of nutri-genetics, which seeks to gain new insights into what genetic factors underlie an individual's dietary preferences and how these factors, if any exist, interact at both cellular and molecular levels. An individual's dietary behaviour could play a major role in susceptibility to obesity, coronary heart disease and metabolic disorders, including type 2 diabetes (Asif, 2014). Therefore, scanning the human genome to identify variants that may have a role in what humans prefer to eat and why, is of medical as well as academic interest. Approaches such as, GWAS, have been used to identify genetic variants that influence preferences for certain foods or beverages. For instance, a GWAS analysis has reported that an olfactory receptor gene (*OR6A2*) influences cilantro (also called coriander, an annual herb used in different ethnic cuisines) preference (Eriksson et al., 2012). Genetic factors have also been reported that influence what we prefer to drink, for instance, liking white wine could be linked to polymorphism in the *HLA-DOA* gene (Pallister et al., 2015). The major histocompatibility complex locus is highly polymorphic and studies in the past have suggested that MHC genotypes could be linked to mate choice and recognition as individuals with different HLA genotype have different bacterial flora affecting body odour (Milinski & Wedekind, 2001; Ober et al., 1997; Pirastu et al., 2015). Based on this, it was proposed that, by regulating the MHC class II molecules, HLA-DOA may have a role in perception of volatile compounds present in white wine which are produced by the bacteria used in wine production (Pirastu et al., 2015). In addition,

studies in monozygotic (MZ) and dizygotic (DZ) twins have reported that dietary preferences have a genetic basis (Hasselbalch et al., 2008). It is therefore likely that there may be some genetic factors which have not yet been identified underlying what humans prefer to eat and drink.

The food items selected to study an individual's preferences cover a wide-range of ingredients and taste stimuli which could play a major role in shaping an individual's health. For example, the high amount of fatty acids (triglyceride) in butter gives it a distinctive flavour while primary alkaloids such as theobromine in dark chocolate gives it a bitter taste (Drewnowski, 2001). On the other hand, vegetables like broccoli may taste bitter to many (Breslin & Spector, 2008) while isothiocyanate causes the burning flavour of the herb watercress (Herr & Buchler, 2010). Both glucosinolate (present in broccoli) and isothiocyanate are thought to have chemo-protective properties (Shapiro et al., 2001). Marmite is known to have a distinctive bitter/salty/umami taste. In conclusion, the different foods selected for the current study were done so as they cover the major categories of tastes such as bitter, sweet, sour and umami.

The aim of this chapter is: a) to investigate the specific genetic factors which influence our preference for butter, soy sauce, parmesan cheese, Marmite, chocolate and dark chocolate (70% cocoa) and b) to examine the genetic causes of variation in the perception of watercress (pepperiness) and broccoli (bitterness) as some people taste watercress as peppery and other as sweet.

5.2 Data collection

Phenotype data were collected, by questionnaire, for butter, soy sauce, parmesan cheese, Marmite, broccoli, watercress, chocolate and dark chocolate (containing 70% cocoa). Participants were asked to rate their preferences for food items on a scale of 1 to 7, where 1 represents 'like greatly' and 7 represents 'dislike greatly' (Figure 5.1, Appendix A.1). The scale for food preferences used in the current study was based on a literature survey (Jones L.V., 1955) and previous GWAS studies have used similar scales (Jones L.V., 1955; Pirastu et al., 2016; Pirastu et al., 2015). For broccoli and watercress, participants were asked to rate bitterness and pepperiness respectively, on the same scale. For chocolate and dark chocolate, participants were asked to simply state if they were likers or dislikers. Participants were shown the actual food item at the time of data collection to ensure that the subject of the question was clear. A summary of the samples used in the current analysis for each phenotype is given in Table 5.1 and Figure 5.1. Participants who had never tried the food could not express their preference, so were excluded from the analysis. For each phenotype analysed, participants who responded as 'Neither like nor dislike' or 'Neither sweet nor bitter' for broccoli and 'Neither sweet nor peppery' for watercress were excluded. This left two groups of individuals for each phenotype: those rating the question 1-3 and those rating it 5-7, which were used to perform a case-control analysis.

Table 5.1 | Characteristics of the samples used in the food preference GWAS.

Phenotype	***Participants (Male/Female)***	***Mean age (SD)***
Butter preference	1593 (752/841)	65.57 (11.66)
Broccoli bitterness	1547 (719/828)	65.58 (11.63)
Soy sauce preference	1412 (644/768)	64.75 (11.48)
Parmesan preference	1443 (658/785)	65.31 (11.55)
Marmite preference	1467 (683/784)	65.55 (11.41)
Watercress pepperiness	1359 (615/744)	65.50 (11.57)
Chocolate preference	1585 (747/838)	65.53 (11.64)
Dark chocolate preference	1567 (735/832)	65.45 (11.67)

Figure 5.1 | Bar plots showing the distribution of the number of participants along with their food preference ratings. A scale from 1 to 7 was used to rate an individual's food preference where 1 denotes 'like greatly' and 7 denotes 'dislike greatly'.

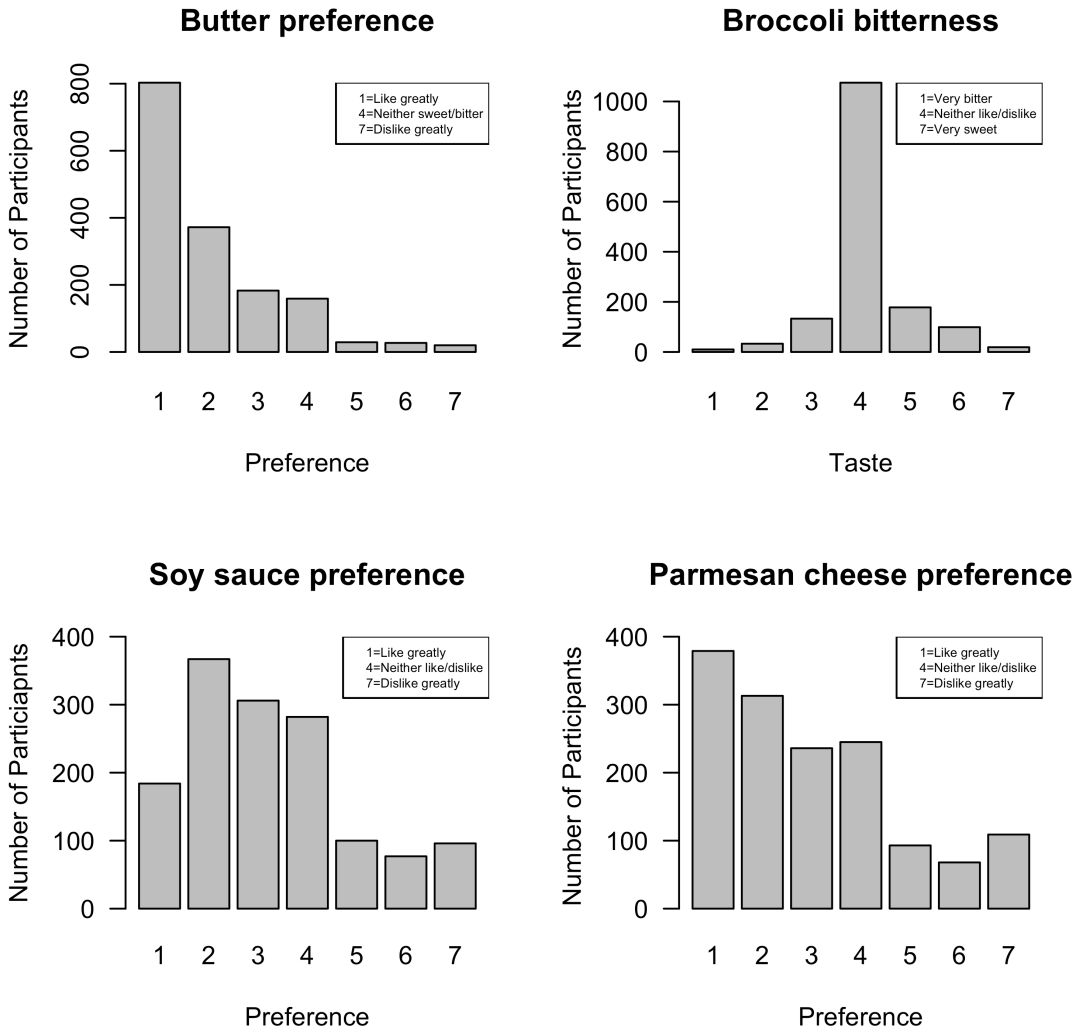
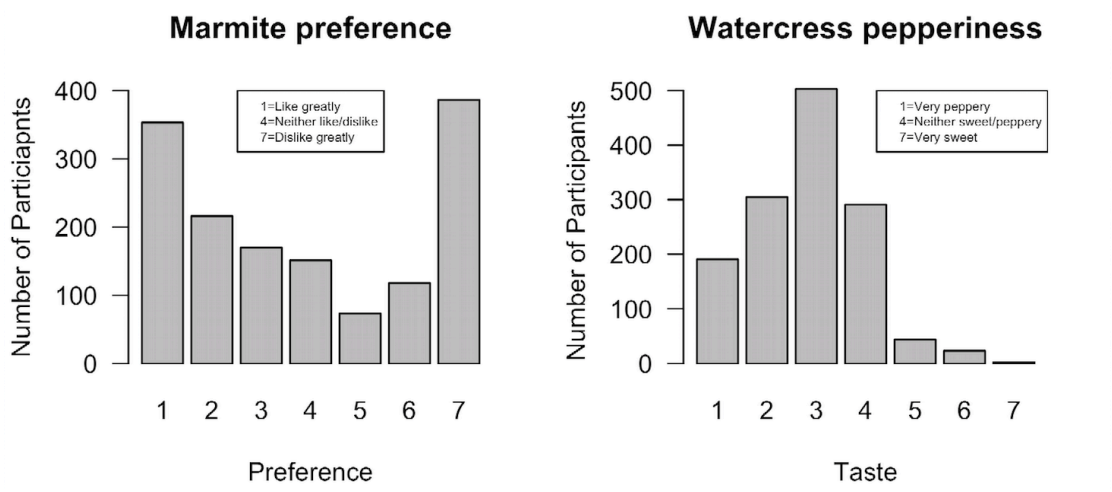


Figure 5.1 | Cont.



5.3 GWAS of food preference

Association analyses of the above-mentioned phenotypes were performed on 519,188 SNPs using Fishers' exact test (3 x 2 genotypic test). The multi-panel Figure 5.3 shows the Manhattan plots of the observed associations for all the phenotypes analysed. These were defined as follows -

Butter preference

For butter preference, 1,358 subjects who rated their preference as 1-3 were tested against 76 subjects who rated their preference as 5-7. Two SNPs reached suggestive genome-wide significance ($P \leq 1 \times 10^{-5}$) as shown in Figure 5.3 (Appendix D.1). The minor allele of the most significant SNP (rs11126303, $P = 2.3 \times 10^{-6,}$ $OR_{Dom} = 5.65$) was positively associated with disliking butter (Figure 5.2, Figure 5.3, Figure 5.4 and Table 5.2). rs11126303 is an intron variant of *KIF3C* on chromosome 2 (Appendix D.2). *KIF3C*, preferentially expressed in the brain, belongs to the kinesin

superfamily whose role is prominent in the intracellular transport in neurons and in the development of cilia in vertebrates (Sardella et al., 1998). Studies have shown that motile cilia of human respiratory epithelia express bitter taste receptors which are identical to the ones found in the nose and tongue (Shah et al., 2009). In addition to rs11126303 on chromosome 2, an intergenic variant (rs10976860, $P = 8.3 \times 10^{-6}$, $OR_{Rec} = 3.12$) within 100 kb of the *PTPRD* region on chromosome 9 had the minor allele positively associated with disliking butter. Protein tyrosine phosphatase, receptor type D (PTPRD) is a member of the protein tyrosine phosphatase (PTP) family of transmembrane proteins that regulate cell signalling and axonal regeneration.

A genome-wide gene expression analysis has shown expression of the *PTPRD* in the circumvallate of primate taste buds and suggested the possibility that it may have a role in downstream signalling of a taste stimulus (Hevezi et al., 2009). Additionally, GWAS studies have also reported that *PTPRD* polymorphism is associated with antipsychotic-induced weight gain (AIWG, antipsychotics are a class of medicines used to treat several disorders such as schizophrenia and often result in weight gain) (Yu et al., 2016) and type 2 diabetes in Han Chinese (Tsai et al., 2010). Although there is no direct functional evidence that *KIF3C* or *PTPRD* have any role in taste perception, considering their tissue-specific expression and their role in different cellular processes makes them good candidates for involvement with food preference.

Figure 5.2 | Volcano plot for the GWAS of butter preference. The observed $-log_{10}$ transformed p-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).

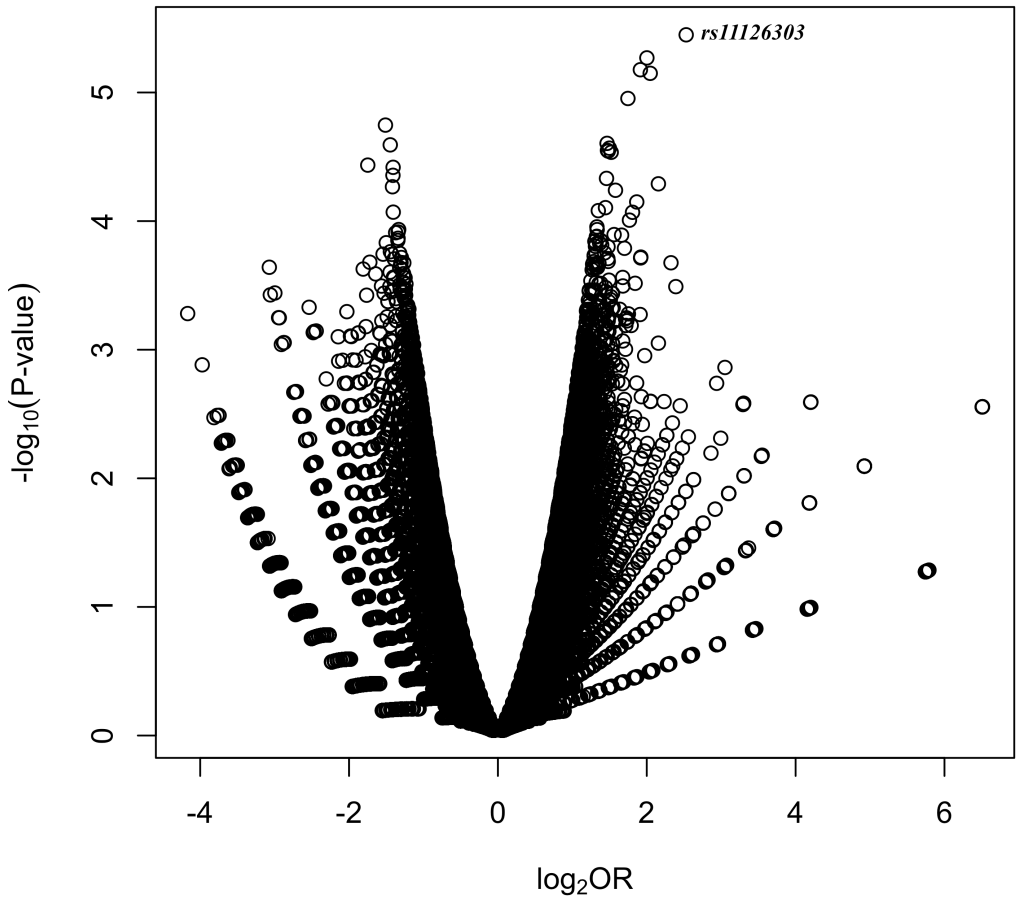


Table 5.2 | Distribution of rs11126303 genotypes in participants with butter preference.

Genotype	***AA***	***Aa***	***aa***	***Total***
Disliking	1	13	62	76
Liking	1	51	1306	1358
Total	2	64	1368	1434

$P_{Dom} = 3.5 \times 10^{-6}$, $OR_{Dom} = 5.65$; $P_{Rec} > 0.05$

Broccoli bitterness

176 subjects who gave a rating of 1-3 (bitter) were tested against 296 subjects who gave a rating of 5-7 (sweet). A total of five SNPs attained suggestive genome-wide significance ($P \leq 1 \times 10^{-5}$) for association with bitter taste perception of broccoli (Figure 5.3, Figure 5.4 and Appendix D.1). The most significant SNP, rs4442431 ($P = 2.0 \times 10^{-6}$, $OR_{het} = 2.69$, $P_{het} = 4 \times 10^{-7}$, Table 5.3) for which the heterozygote was positively associated with broccoli sweetness which indicates possible balancing selection due to heterozygous advantage (Table 5.3). rs4442431 is a regulatory region variant and lies within 300 kb of solute carrier family 35 member F3 (*SLC35F3*) and potassium-channel subfamily K member 1 (*KCNK1*) on chromosome 1 (Appendix D.2). *SLC35F3* is a thiamine transporter and *KCNK1* was suggested to be involved in downstream signal transduction in sour taste perception (Zhang et al., 2014). *KCNK1* (also called *TWIK*-1) is expressed highly in taste buds and regulates excitability of the taste cells (Richter et al., 2004). Another SNP, which attained suggestive genome-wide significance, rs1519113 ($P = 7.6 \times 10^{-6}$, $OR_{Rec} = 5.82$, minor allele positively associated with broccoli sweetness, Table 5.4) lies on chromosome 2 within 100 kb of Meis Homeobox 1 (*MEIS1*). Again, *MEIS1*

has no direct role in the biology of taste buds, but it has been shown to express highly in the lateral olfactory epithelium (OE) and it has been suggested that Meis-expressing cells are stem cells, which mediate neurogenesis of olfactory receptor neurons (Tucker et al., 2010). Association statistics for SNPs that reached suggestive genome-wide significance are shown in Appendix D.1.

Table 5.3 | Distribution of rs4442431 genotypes in participants of broccoli bitterness perception.

Genotype	***AA***	***Aa***	***aa***	***Total***
Bitter	57	62	57	176
Sweet	60	176	60	296
Total	117	238	117	472

$P_{Dom} = 0.004079$, $OR_{Dom} = 1.88$; $P_{Rec} = 0.004079$, $OR_{Rec} = 1.88$

Table 5.4 | Distribution of rs1519113 genotypes in participants of broccoli bitter taste perception GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Bitter	19	40	117	176
Sweet	6	110	180	296
Total	25	150	297	472

$P_{Dom} > 0.05$; $P_{Rec} = 6.7 \times 10^{-5}$, $OR_{Rec} = 5.82$

Soy sauce preference

For soy sauce preference, 857 subjects who rated their preference as 1-3 were tested against 273 subjects who rated their preference as 5-7. A total of five SNPs reached suggestive genome-wide significance (Figure 5.3, Figure 5.4 and Appendix D.1). The most significant SNP, rs12101368 ($P = 9.7 \times 10^{-7}$, $OR_{Rec} = 2.56$, minor allele positively associated with disliking soy sauce) is a non-coding transcript variant near (< 200 kb) *ADAMTS17* (ADAMTS, a disintegrin and metalloproteinase with thrombospondin motifs protein family), which has no known role in taste perception (Appendix D.3, regional association plot). In addition, a single SNP rs4822135 ($P = 1.4 \times 10^{-6}$, $OR_{Rec} = 2.23$) on chromosome 22, two SNPs (rs1531846, $P = 2.4 \times 10^{-6}$, $OR_{Dom} = 2.12$, rs11779424, $P = 5.2 \times 10^{-6}$, $OR_{Dom} = 2.07$) on chromosome 8 and one SNP (rs386057, $P = 6.6 \times 10^{-6}$, $OR_{Dom} = 1.82$) on chromosome 5 were among the SNPs which reached suggestive genome-wide significance. None of these SNPs or associated genes have a clear biological role in taste perception or olfaction.

Parmesan cheese preference

928 subjects who rated their preference as 1-3 were tested against 270 subjects who rated their preference as 5-7. A total of five SNPs attained suggestive genome-wide significance for parmesan cheese preference (Figure 5.3, Figure 5.4 and Appendix D.1). The most significant association was on chromosome 17 (rs2435200, $P = 3 \times 10^{-6}$, $OR_{Rec} = 2.93$, minor allele positively associated with liking parmesan cheese, Table 5.5 and Appendix D.3 for regional association plot) which is an intron variant in microtubule-associated protein tau (*MAPT*). MAPT protein, expressed highly in neurons in the central nervous system, may have a role in olfactory dysfunction

(Dong et al., 2015). Another SNP which reached suggestive genome-wide significance, rs2184898 ($P = 8 \times 10^{-6}$, $OR_{Rec} = 2.93$), is an intergenic variant lies within 100 kb of empty spiracles homeobox 2 (*EMX2*) on chromosome 10 and experimental studies in mice have shown that *EMX2* stimulates odorant receptor (OR) gene expression (McIntyre et al., 2009).

Additionally, a cluster of SNPs attained borderline suggestive significance on chromosome 1. Among these, rs10493631 ($P = 6.5 \times 10^{-6}$, $OR_{Dom} = 1.84$), an intergenic variant within 300 kb of high-mobility group (HMG) box 1 pseudogene 18 (*HMGB1P18*), was the most significant association and *HMGB1P18* has been shown to be associated with childhood obesity in the Hispanic population (Comuzzie et al., 2012).

Figure 5.3 | Manhattan plots for the GWAS of food preferences a) Butter b) Broccoli c) Soy sauce and d) Parmesan cheese. The observed $-log_{10}$ transformed p-values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). The horizontal blue line indicates the threshold for suggestive genome-wide significance ($P = 1 \times 10^{-5}$).

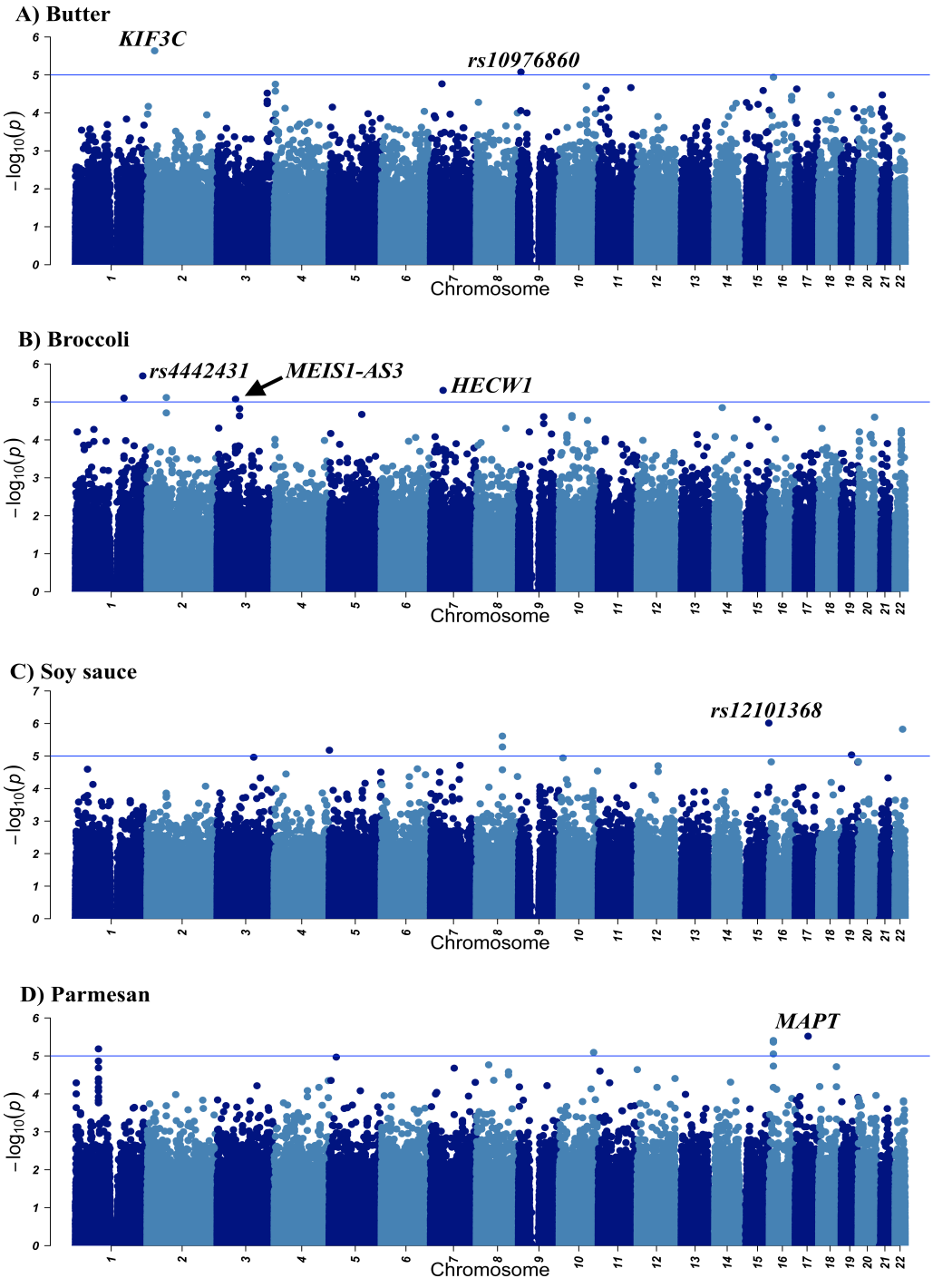


Figure 5.4 | Q-Q plots for the GWAS of food preferences. The $-log_{10}$ of observed p-values (black dots) for each SNP were plotted against the $-log_{10}$ of expected p-values. The red line indicates the null hypothesis of no true association.

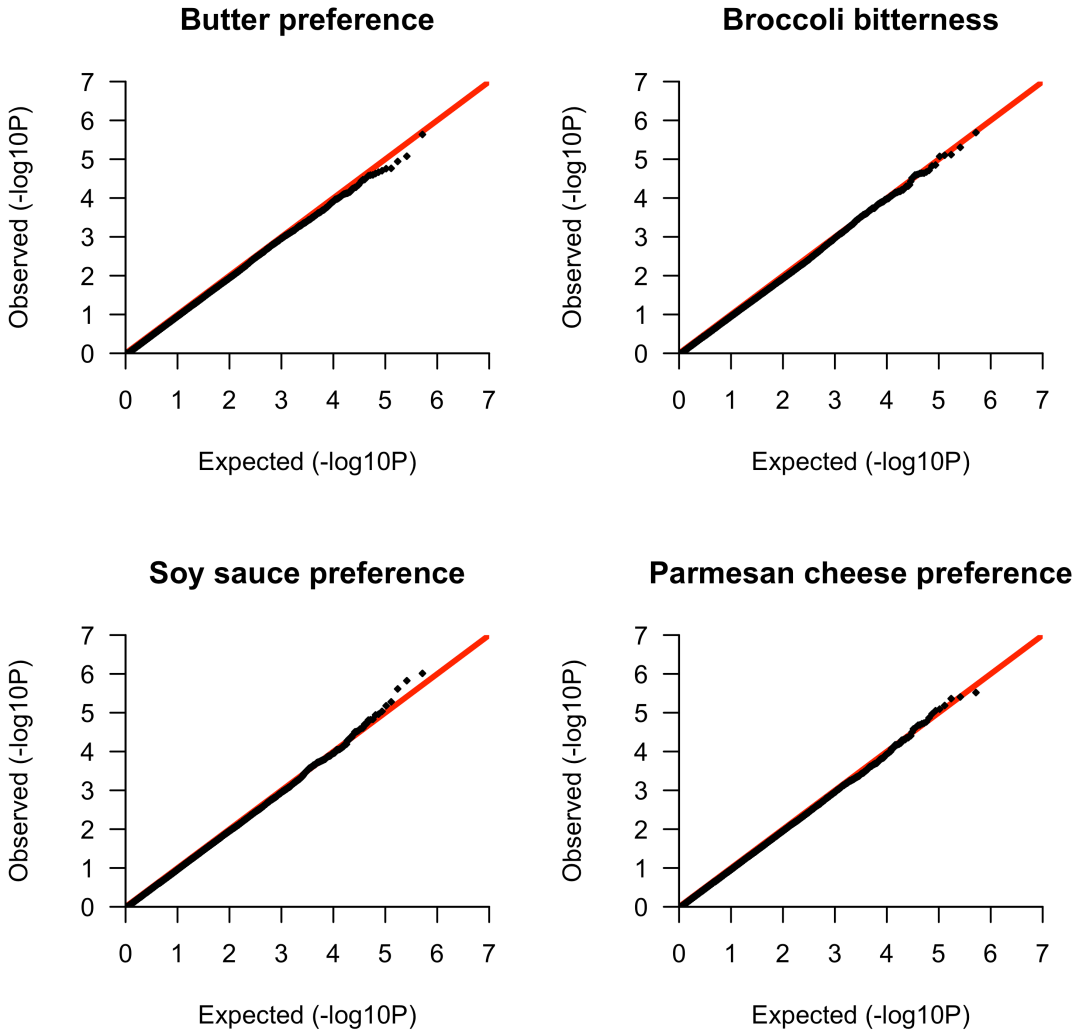


Table 5.5 | Distribution of rs2435200 genotypes in participants with parmesan cheese preference.

Genotype	***AA***	***Aa***	***aa***	***Total***
Disliking	21	134	115	270
Liking	184	429	315	928
Total	205	563	430	1198

$P_{Dom} = 0.01154$, $OR_{Dom} = 1.44$; $P_{Rec} = 9.5 \times 10^{-7}$, $OR_{Rec} = 2.93$

Marmite preference

739 subjects who rated their preference as 1-3 were tested against 577 subjects who rated their preference as 5-7. Three SNPs with strong supporting biological evidence attained suggestive genome-wide significance for Marmite preference (Figure 5.5, Figure 5.6 & Figure 5.7, Appendix D.1). The most significant SNP (rs11979436, $P = 1.2 \times 10^{-6}$, $OR_{Rec} = 10.89$, positively associated with liking Marmite, Table 5.6) located on chromosome 7 lies in membrane-associated guanylate kinase inverted 2 (*MAGI2*), which encompasses ~1.40 Mb. MAGI2 (also known as synaptic scaffolding molecule, or S-SCAM, Appendix D.4), a brain-specific scaffolding protein, interacts with different molecules, including neurotransmitter receptors such as AMPA and NMDA-type glutamate receptors (Deng et al., 2006a). It is important to note here that AMPA and NMDA-type glutamate receptors mediate excitatory synaptic communication in the central nervous system in response to the neurotransmitter glutamate (Bashir et al., 1991). Although Marmite is known to have savoury flavour (Teucher et al., 2007) it also contains monosodium glutamate (MSG) naturally, which is known to elicit a unique

taste called umami (taste of glutamate) and taste cells have been shown to express ionotropic glutamate receptors (iGluRs, for example *NMDA* receptors) which are activated by the neurotransmitter glutamate (Chaudhari et al., 1996). It is also important that metabotropic glutamate receptors (mGluRs, which indirectly activate ion channels) have been shown to function as taste receptors (Chaudhari et al., 2000) while iGluRs have not yet been identified as taste receptors, although some studies in rat taste buds have shown that MSG may regulate iGluRs (Lin & Kinnamon, 1999).

In addition to rs11979436, two other SNPs (rs1045643, $P = 2.5 \times 10^{-6}$, $OR_{Dom} = 1.93$, positive association with liking Marmite, Appendix D.5, and rs2306660 ($P = 3.6 \times 10^{-6}$, $OR_{Rec} = 3.17$, positive association with disliking Marmite), which attained suggestive genome-wide significance, have no suggestive role in taste perception based on the literature survey so could suggest novel variants. Importantly, rs1045643 lies only ~1.40 Mb away from two olfactory receptors *OR1F1* and *OR2C1* (SNPs in *OR1F1* and *OR2C1* were not in LD with rs1045643) on chromosome 16, which could suggest that variation in their enhancer elements affect this phenotype.

Figure 5.5 | Volcano plot for the GWAS of Marmite preference. The observed – log_{10} transformed p-values (Fisher's exact test, recessive model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).

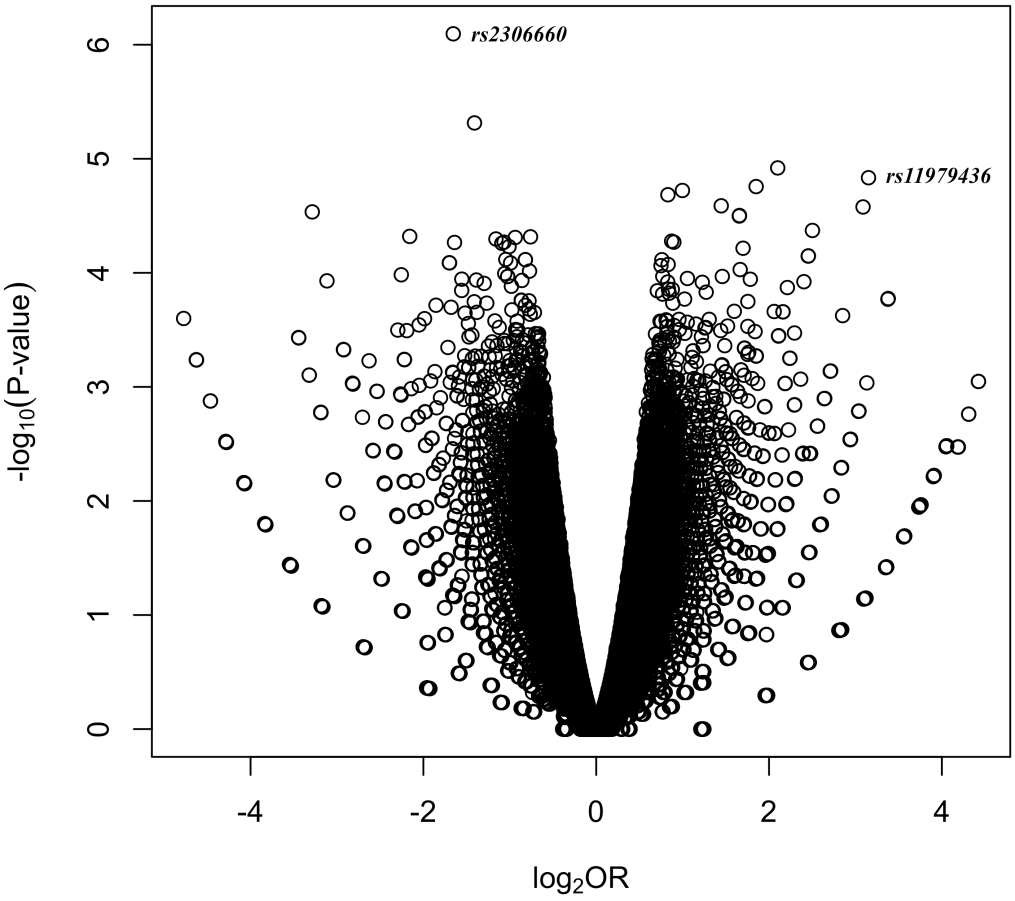


Table 5.6 | Distribution of rs11979436 genotypes in participants with Marmite preference.

Genotype	***AA***	***Aa***	***aa***	***Total***
Disliking	2	105	470	577
Liking	27	176	536	739
Total	29	281	1006	1316

$P_{Dom} = 0.0001456$, $OR_{Dom} = 1.66$; $P_{Rec} = 1.4 \times 10^{-5}$, $OR_{Rec} = 10.89$

Watercress pepperiness

For watercress pepperiness perception, 999 subjects who gave a rating of 1-3 (peppery) were tested against 69 subjects who gave a rating as 5-7 (sweet). Two SNPs on chromosome 5, one on chromosome 7 and one SNP on chromosome 19 showed suggestive genome-wide significance (Figure 5.6, Appendix D.1). The most significant SNP, rs1560160 ($P = 2.2 \times 10^{-6}$, $OR_{Dom} = 5.44$, minor allele positively associated with watercress sweetness, Table 5.7) is a downstream gene variant (*CD180)* on chromosome 5 (Appendix D.4). *CD180* is a cell-surface protein and belongs to the toll-like receptor (TLR) family that, in complex with TLR4, mediates the inflammatory response in B-lymphocytes. Importantly, TLRs have been shown to be expressed in taste buds and their expression in mouse taste epithelium was significantly higher compared to non-gustatory lingual epithelium (Wang et al., 2009). A list of all the SNPs that reached suggestive genome-wide significance is given in Appendix D.1.

Table 5.7 | Distribution of rs1560160 genotypes in participants with watercress pepperiness perception.

Genotype	***AA***	***Aa***	***aa***	***Total***
Sweet	15	47	7	69
Peppery	160	457	380	997
Total	175	504	387	1066

$P_{Dom} = 8.2\times10^{-7}$, $OR_{Dom} = 5.44$; $P_{Rec} = 0.2381$, $OR_{Rec} = 1.45$

Chocolate preference

A total of 1,493 chocolate liking subjects were tested against 92 subjects who said they dislike chocolate. The most significant association was on chromosome 5 (rs818534, $P = 1.2 \times 10^{-7}$, $OR_{Rec} = 4.26$, positive association of minor allele with disliking chocolate, Table 5.8, Appendix D.7), and rs818534 is a regulatory region variant within 100 kb of *SUB1* (also called positive cofactor 4). As shown in the Manhattan plot (Figure 5.6), a cluster of SNPs approaching suggestive genome-wide significance on chromosome 17 was observed. rs17550268 ($P = 2.4 \times 10^{-6}$, $OR_{Dom} = 3.93$, minor allele positively associated with liking chocolate, Appendix D.6) on chromosome 17 lies in the schlafen family member 14 (*SLFN14*). In addition, another variant (rs7661128, $P = 5.4 \times 10^{-6}$, $OR_{Rec} = 9.89$ for liking chocolate) near *GUCY1B3* showed suggestive genome-wide significance. *GUCY1B3* encodes a subunit of the soluble guanylate cyclase (sGC) which mediates conversion of GTP to cGMP, and it has been proposed that sGC may have a role in bitter taste transduction (Rosenzweig et al., 1999).

Table 5.8 | Distribution of rs818534 genotypes in participants with chocolate preference.

Genotype	***AA***	***Aa***	***aa***	***Total***
Dislike	29	31	32	92
Like	145	699	647	1491
Total	174	730	679	1583

$P_{Dom} = 0.1283$, $OR_{Dom} = 1.43$; $P_{Rec} = 2.8 \times 10^{-8}$, $OR_{Rec} = 4.26$

Dark chocolate preference

A total of 1,069 dark chocolate liking subjects were tested against 498 subjects who said they dislike dark chocolate. Three SNPs showed suggestive genome-wide significance for dark chocolate preference (Figure 5.6). Among these, two SNPs (rs969092, $P = 2.4 \times 10^{-6}$, $OR_{Rec} = 3.68$, minor allele positively associated with liking dark chocolate and rs1497177, $P = 3.2 \times 10^{-6}$, $OR_{Rec} = 3.62$, minor allele positively associated with liking chocolate, Table 5.9 and Appendix D.7) lie near *RP11-630C16.2* (lincRNA) on chromosome 12. An intergenic variant, rs7460061, on chromosome 8 ($P = 5.9 \times 10^{-6}$, $OR_{Dom} = 1.44$ minor allele positively associated with disliking dark chocolate) also showed suggestive genome-wide significance, which is not known to have any role in taste perception. The multi-panel Q-Q plots are shown in Figure 5.7.

Table 5.9 | Distribution of rs969092 genotypes in participants with dark chocolate preference.

Genotype	***AA***	***Aa***	***aa***	***Total***
Disliking	10	155	333	498
Liking	75	384	609	1068
Total	85	539	942	1566

$P_{Dom} = 0.0002046$, $OR_{Dom}=1.52$; $P_{Rec} = 1.9\times10^{-5}$, $OR_{Rec} = 3.6$

Figure 5.6 | Manhattan plots for the GWAS of food preferences A) Marmite, B) Watercress, C) Chocolate and D) Dark chocolate. The observed $-log_{10}$ transformed p-values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). The horizontal blue line indicates the threshold for suggestive genome-wide significance ($P = 1 \times 10^{-5}$).

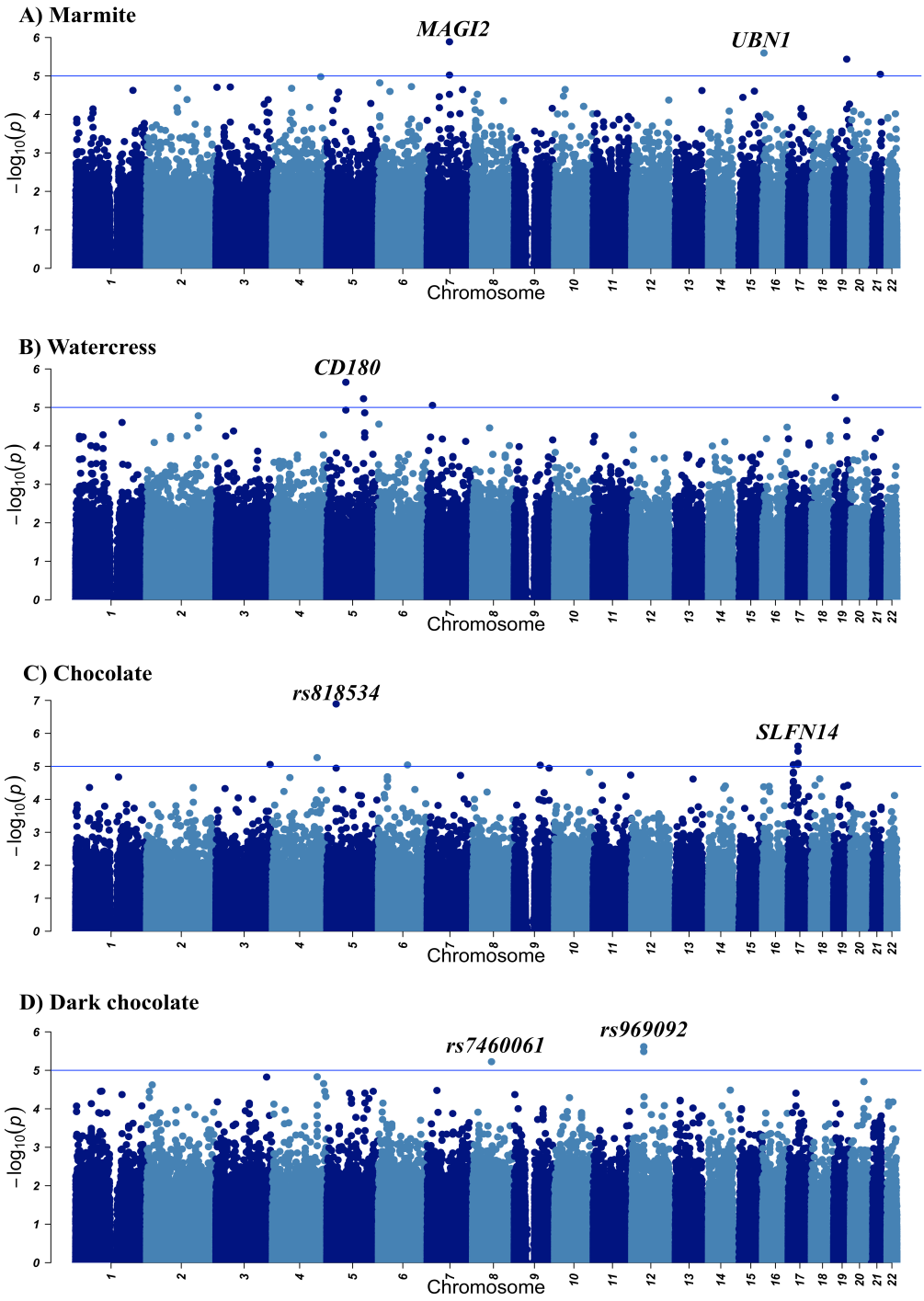
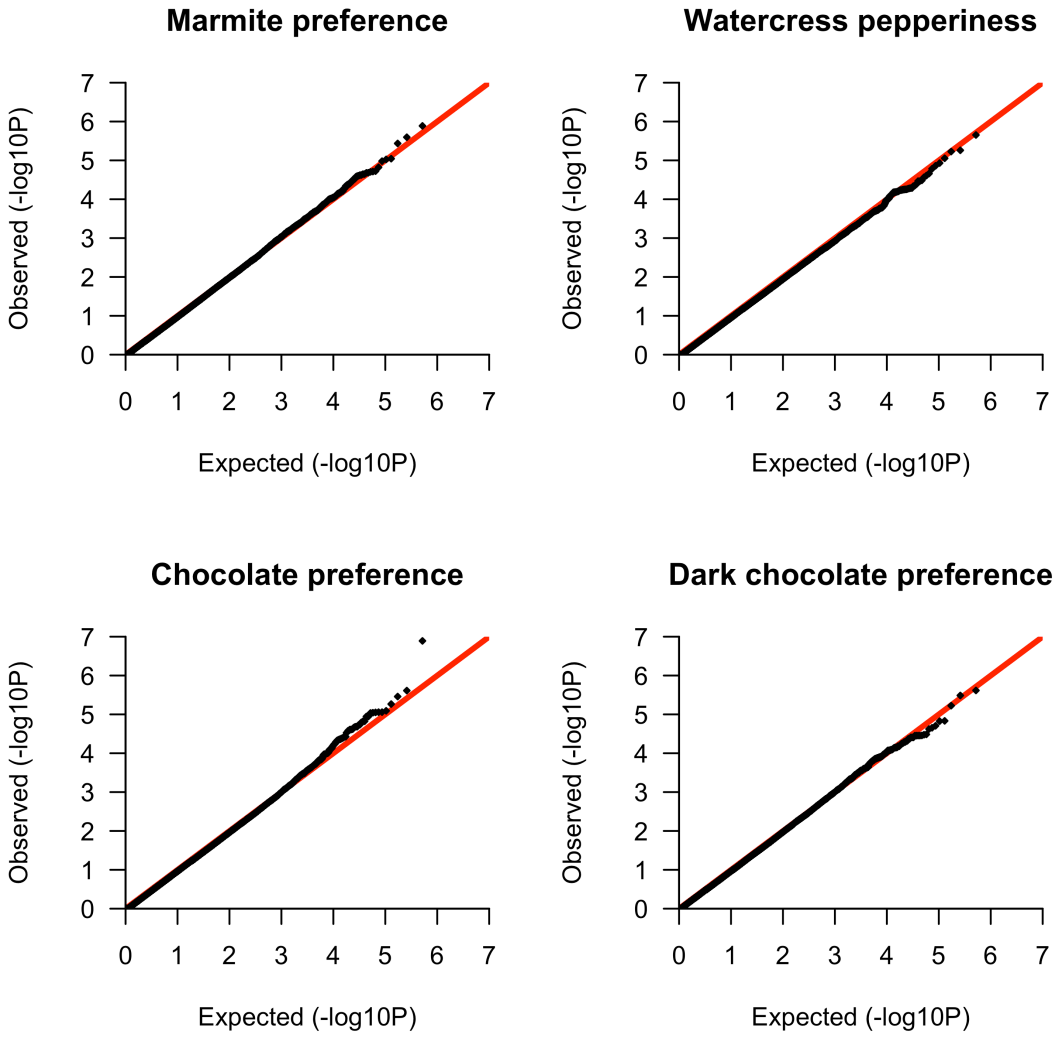


Figure 5.7 | Q-Q plots for the GWAS of food preferences. The $-log_{10}$ of observed p-value*s* (black dots) for each SNP were plotted against the $-log_{10}$ of expected p-values. The red line indicates the null hypothesis of no true association.



5.4 **Discussion**

Previous studies have shown that preference for certain foods is genetically influenced, and understanding these factors would help in dissecting the genetic basis of different diet-related disorders. The current analysis investigates the genetic basis of preference for certain foods in a well-characterised population of the British Isles. A total of thirty-two SNPs attained suggestive genome-wide significance ($P = 1 \times 10^{-5}$). The most significant association (rs11126303, $P = 2.3 \times 10^{-6,}$ $OR_{Dom}= 5.65$) observed for butter taste preference, is an intron variant in the *KIF3C* gene. *KIF3C* is a member of the kinesin superfamily and is preferentially expressed in the brain (Sardella et al., 1998) and experimental studies in mouse have suggested that *KIF3C* may have a role in fast axonal transport (Yang & Goldstein, 1998). Studies have also demonstrated that impairment in axonal transport is associated with taste response dysfunction; therefore, it seems likely that the axonal transport role of *KIF3C* has a role in transduction of taste signals in neurons (Oakley et al., 1981).

An intergenic variant rs10976860 near *PTPRD* also showed suggestive genome-wide significance with butter preference. Studies in the past have linked *PTPRD* with obesity and diabetes (Chen et al., 2015). It is also important that expression enrichment profiling of *PTPRD* showed that *PTPRD* mRNA is preferentially expressed in the frontal cortex (Yu et al., 2016). Given that *PTPRD* has been shown to express in taste buds as well as in olfaction relevant brain regions such as the frontal cortex, a possible role of this protein in influencing taste behaviour cannot be denied.

rs4442431 ($P_{het} = 4 \times 10^{-7}$, $OR_{het} = 2.69$) was positively associated with broccoli sweetness perception which indicates possible balancing selection due to heterozygous advantage. It could be the case that tasting broccoli as sweet is of particular advantage. rs4442431 is located in a regulatory region near *KCNK1* and *SLC35F3*. *KCNK1* is a member of the potassium channel superfamily and is preferentially expressed in taste buds. Experimental studies in mice have shown that *TWIK*-1 (mouse homologue of *KCNK1*) is highly expressed in taste buds and may have a role in taste perception (Richter et al., 2004). Interestingly, ion-channel-derived polarization and depolarization is a common mechanism underlying signal transduction in taste perception (Chapter 1) and *KCNK1* is a strong candidate gene for mediating polarity of the taste cells.

An intron variant (rs2435200, $P = 3 \times 10^{-6}$, $OR_{Rec} = 2.93$) in *MAPT* on chromosome 17 showed suggestive genome-wide association with parmesan cheese preference. *MAPT (*microtubule-associated protein tau) is a major susceptibility gene for Parkinson's disease and a genome-wide meta-analysis involving gene-based and pathway-enrichment analyses proposed that this locus underlies poor sense of smell (Dong et al., 2015). In addition to the *MAPT* region, an intergenic variant (rs10493631, $P = 6.5 \times 10^{-6}$, $OR_{Dom} = 1.84$) attained suggestive genome-wide significance and a cluster of SNPs supported this observation on chromosome 1. This variant, which lies near *HMGB1P18* (a pseudogene), has been shown to be associated with susceptibility to obesity in a genome-wide analysis in the Hispanic population (Comuzzie et al., 2012*).* It is possible that polymorphism in rs10493631 is related to obesity risk via influencing individuals' food preferences.

An intron variant (rs11979436, P =1.2 × 10^{-6}, OR_{Rec} = 10.89) in *MAGI2* on chromosome 7 showed an association with Marmite preference. MAGI2 (also called synaptic scaffolding molecule, or *S-SCAM*) is preferentially expressed in the brain and interacts with a wide variety of molecules including hyperpolarization-activated cation channels, and glutamate (AMPA and NMDA-type receptors) receptors. Interestingly, both hyperpolarization-activated cation channels and AMPA and NMDA-type receptors play fundamental roles in response to different taste stimuli (Stevens et al., 2001). Moreover, AMPA and NMDA are both types of glutamate receptors, which mediate synaptic signalling in the central nervous system (Deng et al., 2006a). Marmite contains monosodium glutamate and thus it seems likely that polymorphism in *MAGI2,* given the large effect size, strongly influences an individual's preference for Marmite.

5.5 **Conclusions**

This study investigates the genetic basis of food preference in a well-characterised British population and highlighted some novel variants, which could be the genetic factors that influence dietary preferences. The current analysis reports *KIF3C, PTPRD, KCNK1* and *MAGI2* as novel candidate for taste perception (based on suggestive p-values and relatively large odds ratios observed in the current analysis) within taste buds and further experimental studies are required to validate these findings.

6 Genetics of odour perception

6.1 Introduction

Humans can sense and differentiate between thousands of different chemical compounds. The ability to sense these is mediated by olfactory receptors found in the olfactory epithelium (OE) of the nose, where a particular ligand (odour) associates with olfactory receptor (Buck & Axel, 1991). Olfactory receptors (OR) belong to the G-protein coupled receptors (GPCR) family. The olfactory receptor gene family is one of the biggest gene families in the human genome having ~636 members (339 olfactory receptor genes and 297 olfactory receptor pseudo genes) distributed over ~ 50 different loci on 21 chromosomes (Malnic et al., 2004). Proteins or combinations of proteins encoded by these genes mediate how an individual perceives the odours of different chemicals. Perception of odour plays an important role in day-to-day life, from selection of food to identifying potentially hazardous situations in one's surroundings, and also in social behaviour.

Since the molecular basis of ORs and their signal transduction mechanisms have been elucidated (Buck & Axel, 1991), and with the advancement in genotyping technologies and GWAS, our understanding of olfactory systems has increased vastly. For example, a recent study has shown that cilantro preference (soapy taste perception of cilantro) is influenced by a variant near an OR gene (*OR6A2*) (Eriksson et al., 2012) while another study reports that the ability to smell the urinary metabolites of asparagus is influenced by a genetic variant near *OR2M7* (Eriksson et al., 2010b). It has also been shown that loss or poor sense of smell is the characteristic of most neurodegenerative and neurodevelopmental disorders like autism spectrum disorder

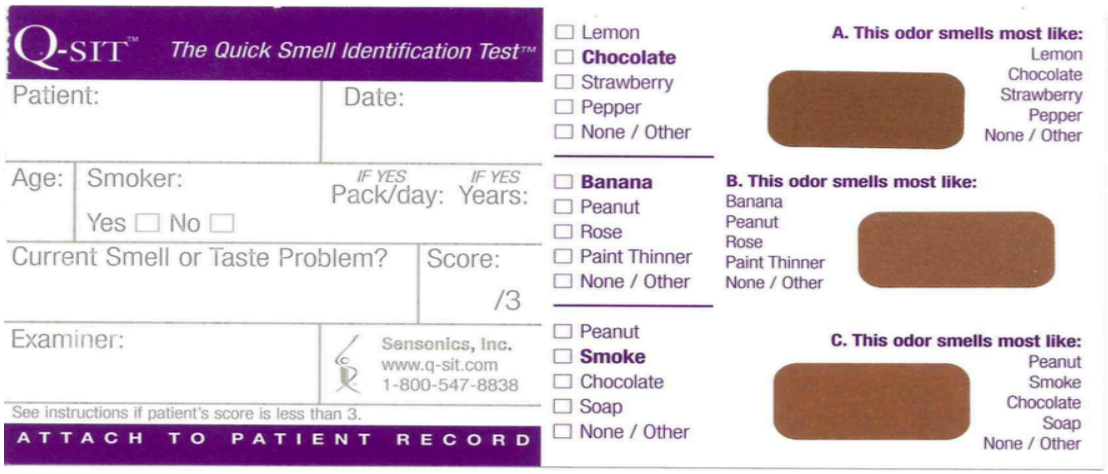
(Rozenkrantz et al., 2015), Parkinson's and Alzheimer's diseases. Therefore, in addition to its role in influencing food selection and social behaviour, odour perception could potentially be helpful in early diagnosis of neurodegenerative disorders like Parkinson's and Alzheimer's diseases (Dong et al., 2015). The current research investigates the genetic basis of odour perception (chocolate, banana and smoke, chosen based on the available validated method to assess the sense of smell) in a well-characterised population of the British Isles.

6.2 Phenotype data collection

6.2.1 Smell identification test

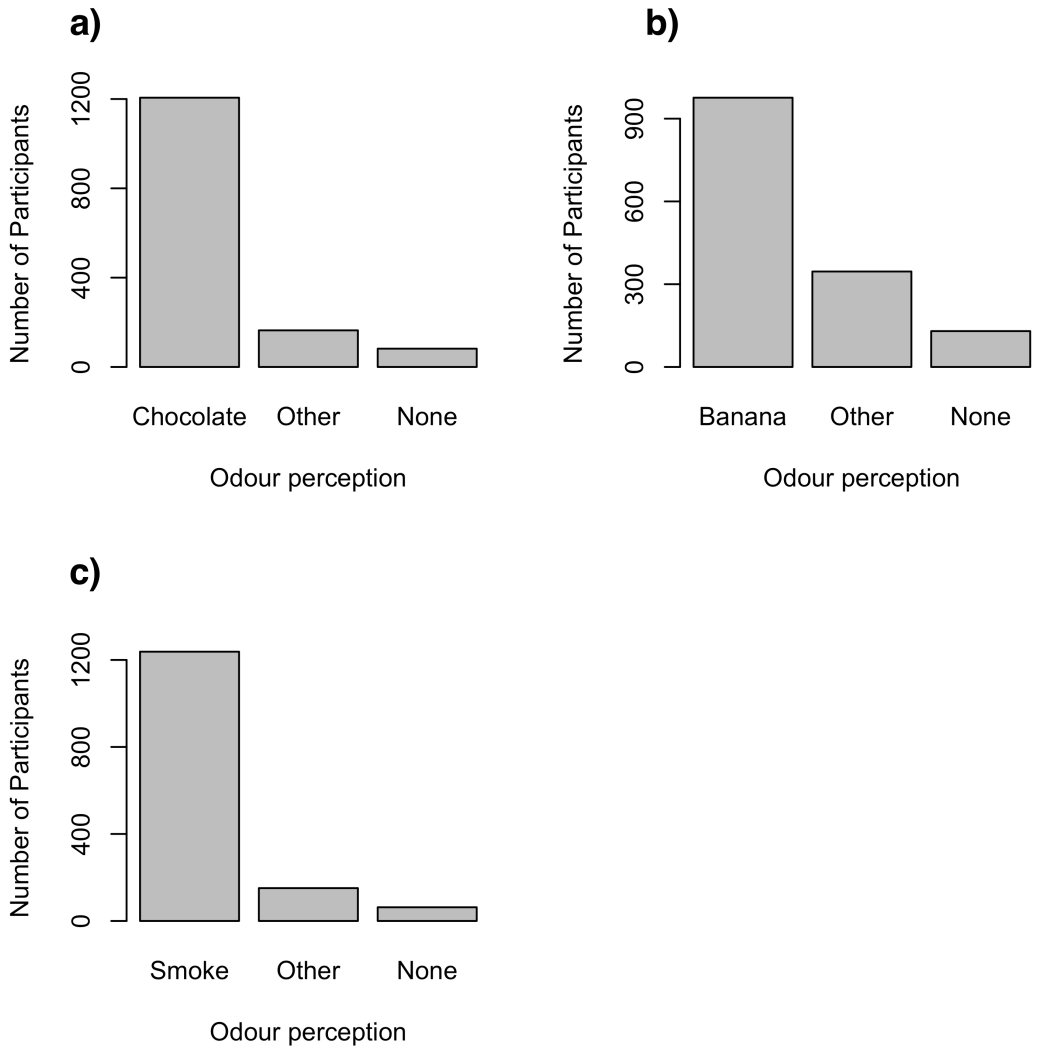
Each participant was provided with the Quick Smell Identification Test™ (Q-SIT, Sensonics international) which is a validated method to assess the sense of smell, and is highly sensitive in identifying olfactory loss (Jackman & Doty, 2005). The Quick Smell Identification Test™ contains three different odorants, which resemble day-to-day life odorants found in chocolate, banana and smoke. Each participant was asked to smell each odorant one at a time and to report what they thought it was from five possible options as shown in the Figure 6.1. Participants were also asked to report their age, smoking status and any medical conditions which could influence olfactory perception.

Figure 6.1 | The Quick Smell Identification Test.



After performing standard SNP quality-control procedures on the PoBI dataset (Chapter 2), 1,576 samples had genotype and phenotype data available for odour perception. Subjects who were suffering from a cold at the time of sample collection (n = 124) were excluded from further analysis, as any smell inability they demonstrated was likely to be non-genetic. A total of 1,452 samples were considered for further analysis. The multi-panel Figure 6.2 shows the number of individuals and their recorded response.

Figure 6.2 | Multi-panel bar plots showing the odour perception data: a) chocolate, b) banana and c) smoke.



6.3 **GWAS of chocolate odour perception**

Genome-wide association analysis was performed using Fisher's exact test (3 x 2 genotypic test) in a total of 1,370 participants (1,206 perceived the smell of chocolate, 164 responded as 'Other', and 82 with the response 'None' were excluded) to scan for genetic factors underlying the ability to detect the chocolate odour. Although no SNP reached genome-wide significance ($P \leq 5 \times 10^{-8}$), a SNP on chromosome 13,

rs10507739, showed suggestive genome-wide significance (Figure 6.3, Figure 6.4 and Table 6.1). rs10507739 ($P = 4.6 \times 10^{-7}$, $OR_{Dom} = 2.34$ minor allele positively associated with smelling chocolate, Table 6.2) on chromosome 13 (q21.32) is an intron variant in *PCDH9* (protocadherin 9). *PCDH9* is a member of the cadherin superfamily, largely expressed in the central nervous system, which mediates calcium dependent cell-cell recognition (Sano et al., 1993). Interestingly, a SNP near *PCDH9* (rs17252438) had been shown to have suggestive genome-wide significance in a meta-analysis (in which phenotype data were collected using the 12-item Brief Smell Identification Test, including chocolate as one of the odorants) of the sense of smell (Dong et al., 2015) but it was not significant ($P = 0.8416$, $OR_{Dom} = 1.04$) in the current analysis. rs10507739 is located ~400 kb from rs17252438 and showed effectively no linkage disequilibrium ($r^2 = 0.0003$) with it. The Q-Q plot is shown in Appendix E.1.

In addition to the SNPs on chromosome 13, rs11632280 ($P = 1.5 \times 10^{-6}$, $OR_{Dom} = 1.87$, Appendix E.2) on chromosome 15 showed suggestive association with smelling chocolate. rs11632280 is an intergenic variant lying within 300 kb of *PWRN4* and *NDN* (also known as Necdin), and *NDN* is known to be involved in neurodevelopmental disorders and control of sensory neurons during embryonic development (Andrieu et al., 2006). A cluster of olfactory receptors on chromosome 15 (q11.2) lies near *NDN* (Figure 6.5). This cluster of olfactory receptors which is within ~2 Mb of rs11632280, implicates two protein coding olfactory genes *(OR4M2 & OR4N4*) and five olfactory receptor pseudogenes *(OR4Q1P, OR11H3P, OR11K1P, OR4H6P, OR4N3P)*. A regional association plot of chromosome 15 (q11) is shown in Figure 6.5.

A SNP (rs6944822, P = 8.3 × 10^{-6}, OR_{Rec} = 1.82, minor allele positively associated with inability to smell chocolate) on chromosome 7 showed suggestive association with odour perception of chocolate. rs6944822 is an intron variant in *KIAA1549* on chromosome 7 (q34). rs6944822 lies within 100 kb of *TMEM213* (transmembrane protein 213) and it could be of particular interest as this gene has been shown to express in the olfactory receptor neurons (Kanageswaran et al., 2015)

Table 6.1 | Association of genetic variants with chocolate odour perception in the PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***p-value[b]***	***OR***	***r^2***
rs10507739	13	67125380	*PCDH9*	4.6×10^{-7}	2.34	-
rs9540832	13	67158351	*PCDH9*	1.3×10^{-5}	1.97	.69
rs11632280	15	24326612	*PWRN4*/NDN*	1.6×10^{-6}	1.87	-
rs6944822	7	138580906	*KIAA1549*	8.3×10^{-6}	1.82	-
rs750270	11	133086298	*OPCML*	1.2×10^{-5}	2.55	-

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared (calculated with rs10507739); [b]Fisher's exact test; OR, Odds ratio; *RNA gene

Table 6.2 | Distribution of rs10507739 genotypes in participants of chocolate odour perception.

Genotype	***AA***	***Aa***	***aa***	***Total***
Not smelling	2	46	116	164
Smelling	96	498	612	1206
Total	98	544	728	1370

$P_{Dom} = 1.1\times10^{-6}$, $OR_{Dom} = 2.34$, $P_{Rec} = 0.0005286$, $OR_{Rec} = 7.00$

Figure 6.3 | Volcano plot for the GWAS of chocolate odour perception. The observed $-log_{10}$ transformed p-values (Fisher's exact test, dominant model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).

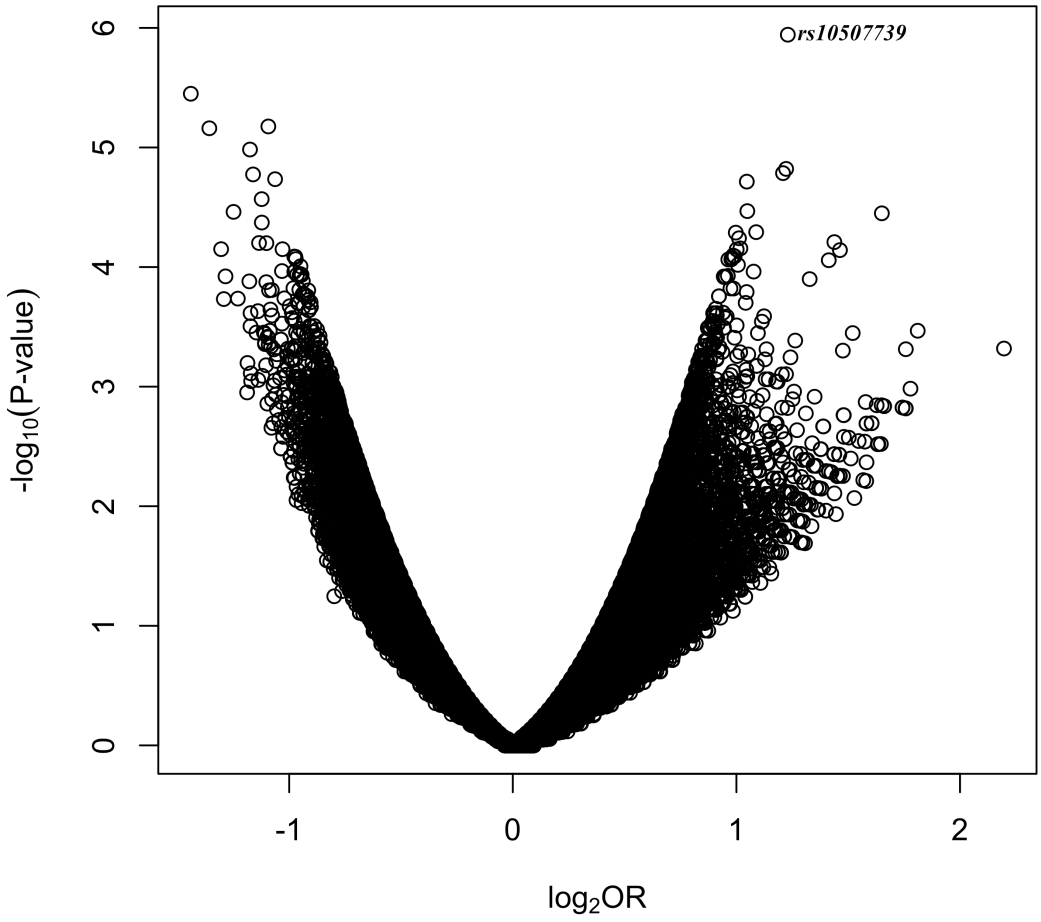


Figure 6.4 | Manhattan plot for the GWAS of chocolate odour perception. The observed $-log_{10}$ transformed p-values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). Each dot represents a SNP on their chromosomal locations. The horizontal blue line indicates the threshold for suggestive genome-wide significance ($P = 1 \times 10^{-5}$).

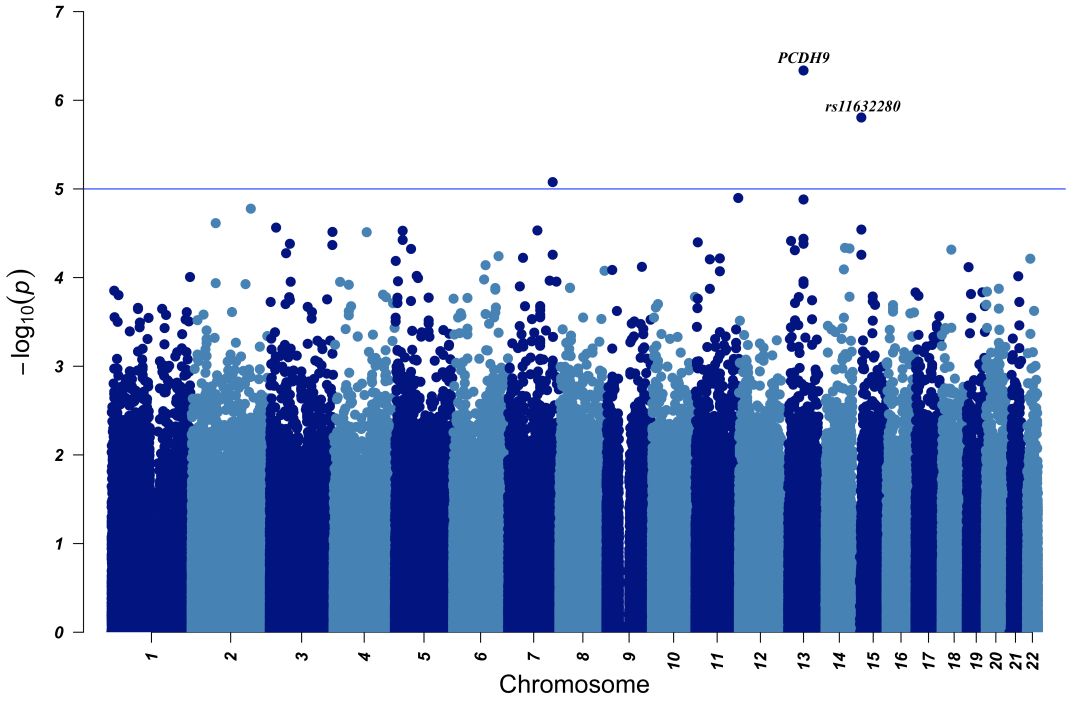
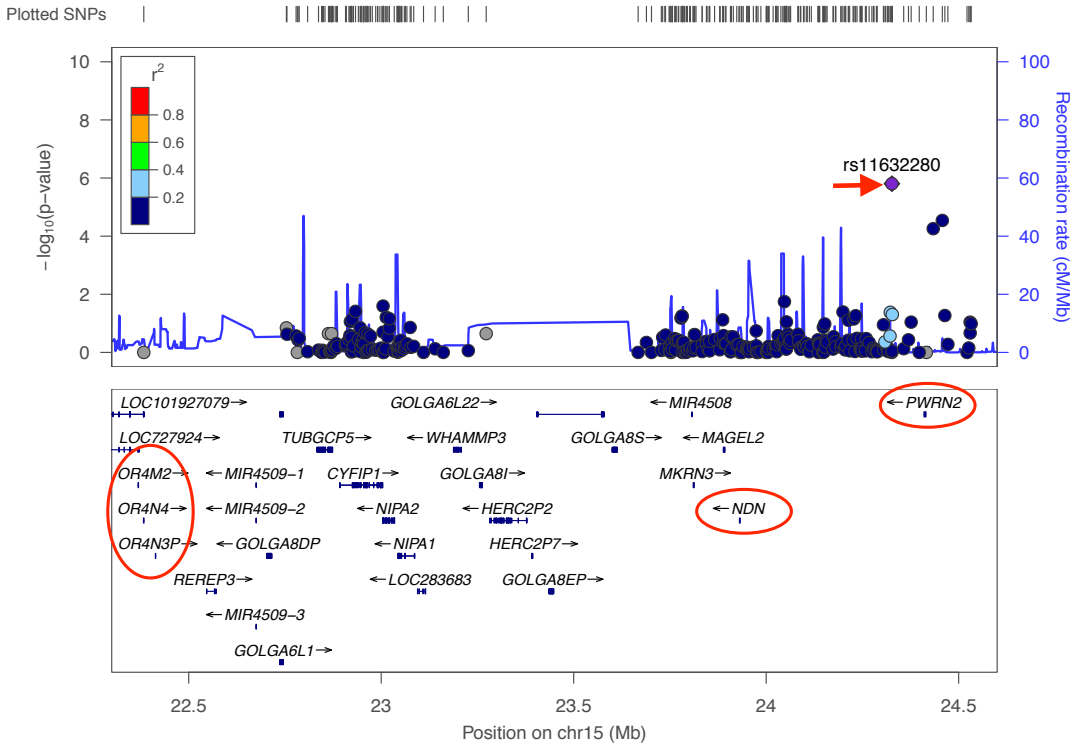


Figure 6.5 | Regional association plot showing rs11632280 on chromosome 15 (q11.2). Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 15. The purple diamond represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.



6.4 Smoke odour perception

For smoke odour perception, genetic associations were tested in 1,389 participants on 519,188 SNPs using Fisher's exact test (3 x 2 genotypic test). Two SNPs, on chromosome 16 (rs7199415) and 17 (rs227722), attained suggestive genome-wide significance in addition to strong supporting biological evidence (Figure 6.6 and Table 6.3). rs7199415 ($P = 3.6 \times 10^{-7}$, $OR_{Rec} = 4.16$, minor allele positively associated with the inability to smell smoke, Table 6.4, Figure 6.7, Appendix E.5) is an

intergenic variant located on chromosome 16 between *CDH8* and *CDH11*. Cadherin-11 is a member of the cadherin superfamily and has been shown to express in the mitral cells of the olfactory bulb in mouse (Akins et al., 2007). Interestingly, mitral cells play a fundamental role in olfaction by transferring information to different parts of the brain, including the piriform cortex (olfactory cortical region in the human brain). rs227722 ($P = 8.6 \times 10^{-7}$, $OR_{Rec} = 2.44$, minor allele positive association with inability to smell, Appendix E.3) is an intergenic variant on chromosome 17 lying between *NOG* and C17orf67 and is ~1.5 Mb away from the protein coding olfactory receptor genes *OR4D1* and *OR4D2*. Interestingly, the SNPs that lie in *NOG* ($P > .05$) and *OR4D2* ($P = 0.056$) cluster did not appear significant. However, none of the SNPs in the *OR4D1, OR4D2* and *NOG* were in LD (for *OR4D2* $r^2 = 0.00014$ and for *NOG*, $r^2 = 0.010$) with rs227722. The molecular function of NOG protein is to modulate the activity of BMP (bone morphogenetic proteins) protein which is known to block the production of neurons in the subventricular zone (the site of neurogenesis) in the central nervous system. Importantly, NOG has been shown to promote neurogenesis of the subventricular cells (Altmuller et al., 2001). Another SNP on chromosome 17 (rs227715, $P = 2.3 \times 10^{-5}$, $OR_{Dom} = 1.64$, minor allele positive association with the ability to smell) showed suggestive genome-wide significance.

A SNP on chromosome 2, rs2287611 ($P = 2.6 \times 10^{-6,}$ $OR_{Dom} = 3.78$, minor allele positive association with the ability to smell), is an intron variant in *DHRS9* showing mild association. Dehydrogenase/reductase SDR family member 9 (*DHRS9*) is a member of a biotransformation enzyme family known to metabolise a variety of chemical compounds including steroid hormones, retinoids etc (Wang et al., 1999). The association of *DHRS9* is of particular interest as a previous study has shown that short chain dehydrogenase/reductase is expressed in the olfactory organs of

Drosophila melanogaster (Wang et al., 1999), which could be an important part of the odorant degradation mechanism.

Another SNP (rs480700, P = 6.4 × 10^{-6}, OR_{Dom} = 1.85) that attained suggestive genome-wide significance lies in *NFIB* (nuclear factor I B) on chromosome 9. *NFIB* has been shown to express in the olfactory bulb (OB) during embryonic and postnatal development in mouse (Plachez et al., 2012). The same study has shown that *NFIB* is expressed in the mitral cells of the olfactory bulb (OB). Importantly, as discussed earlier, mitral cells in the olfactory bulb play a central role in olfaction where they receive information from olfactory receptor neurons and transfer that to different parts of the brain including the piriform cortex.

Table 6.3 | Association of genetic variants with smoke odour perception in the PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***p-value[b]***	***OR***
rs7199415	16	64087122	-	3.6×10^{-7}	4.16
rs227722	17	54779481	-	8.7×10^{-7}	2.44
rs2287611	2	169925861	*DHRS9*	2.6×10^{-6}	3.78
rs480700	9	14079067	*NFIB*	6.4×10^{-6}	1.85

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; [b]Fisher's exact test; OR, Odds ratio;

Table 6.4 | Distribution of rs7199415 genotypes in participants of smoke odour perception.

Genotype	***AA***	***Aa***	***aa***	***Total***
Not smelling	28	41	82	151
Smelling	64	377	796	1237
Total	92	418	878	1388

$P_{Dom} = 0.01$, $OR_{Dom} = 1.51$, $P_{Rec} = 7.9 \times 10^{-8}$, $OR_{Rec} = 4.16$

Figure 6.6 | Manhattan plot for the GWAS of smoke odour perception. The observed $-log_{10}$ transformed -values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). Each dot represents a SNP on their chromosomal location. The horizontal blue line indicates the threshold for suggestive genome-wide significance ($P = 1 \times 10^{-5}$).

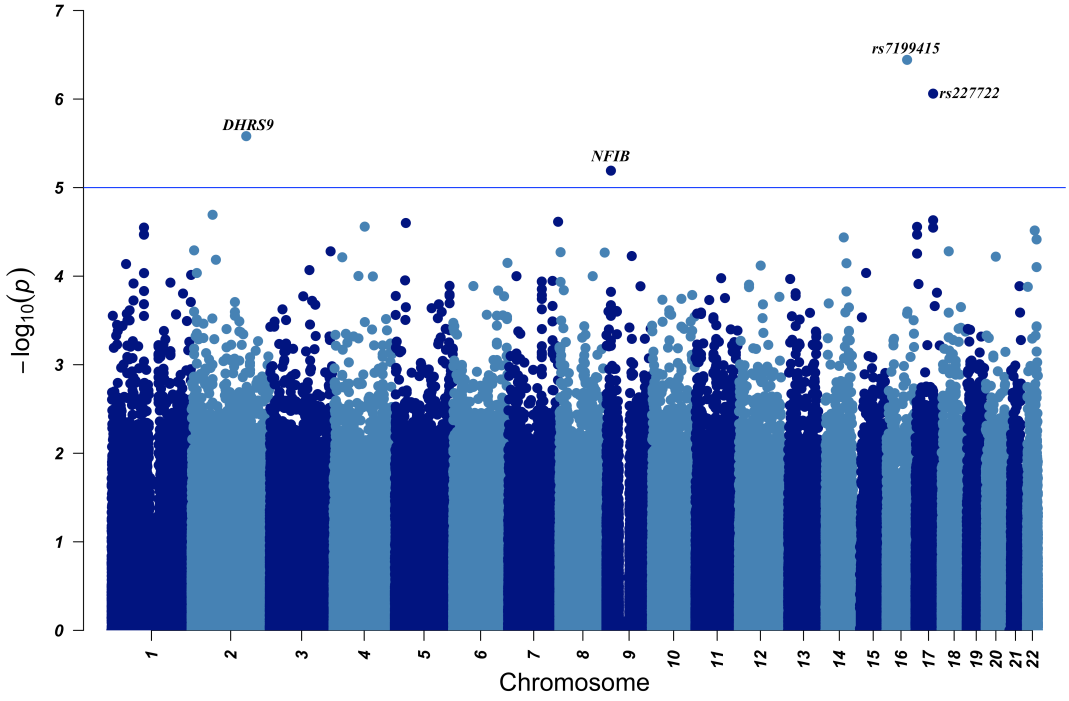
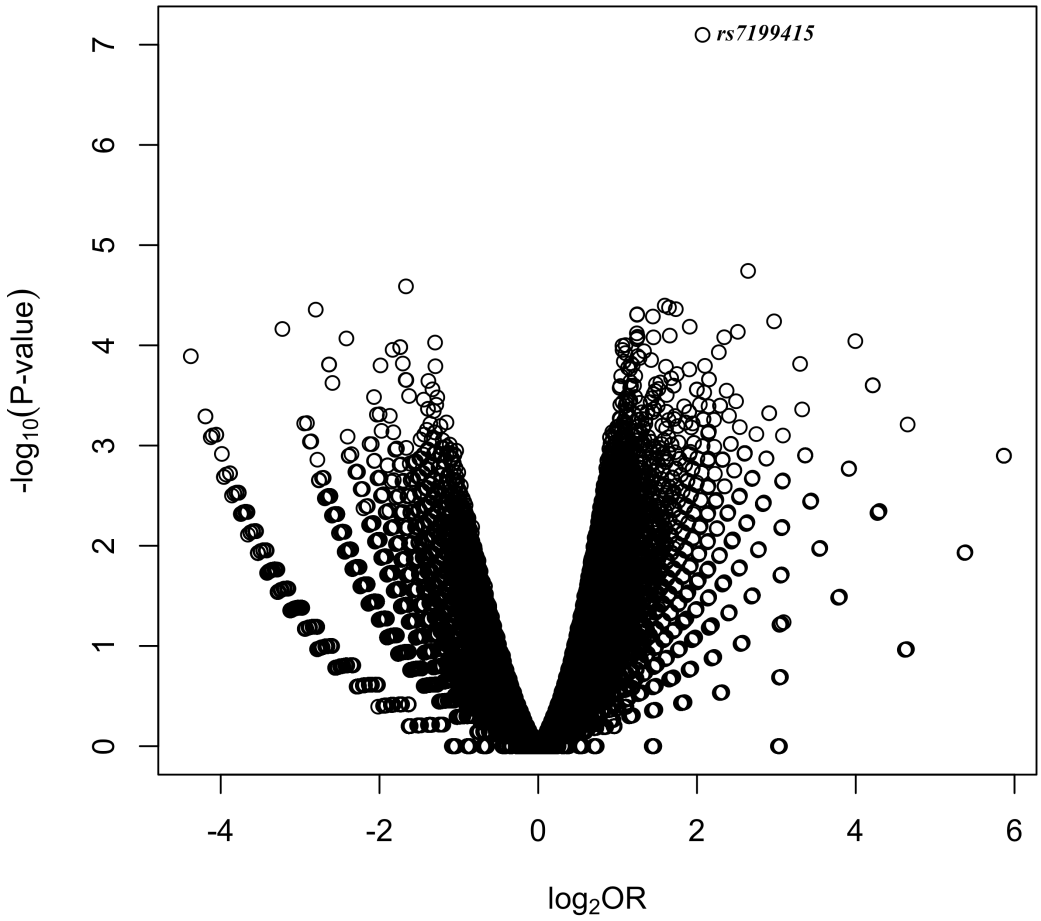


Figure 6.7 | Volcano plot for the GWAS of smoke odour perception. The observed – log_{10} transformed p-values (Fisher's exact test, recessive model) are plotted (y axis) against $\log_2$ of odds ratios (x axis).



6.5 **Banana odour perception**

To scan for genetic variants underlying sensory perception of banana smell, 1,322 individuals (976 perceived the smell of banana and 346 perceived the smell as 'other') were tested using Fisher's exact test (3 x 2 genotypic test). Although no SNP reached genome-wide significance, four SNPs on chromosome 10 showed suggestive genome-wide significance (Figure 6.8, Table 6.5) with minor alleles positively associated with inability to smell the odour of banana. rs1336115 ($P = 3.6 \times 10^{-7}$, $OR_{Dom} = 1.72$, Table 6.6), rs7895853 ($P = 5.8 \times 10^{-7}$, $OR_{Dom} = 1.65$), rs1932582 ($P = 7.3 \times 10^{-7}$, $OR_{Dom} = 1.68$) and rs10787577 ($P = 1.4 \times 10^{-6}$, $OR_{Dom} = 1.65$) all lie in *ATRNL1* on chromosome 10, and the latter three showed strong linkage disequilibrium with the top associated SNP, rs1336115 (Table 6.5 and Appendix E.5). *ATRNL1* (Attractin-like 1 protein) is a single-pass membrane protein, which interacts with melanocortin-4 receptor and regulates energy homeostasis, and has been shown to express in the mouse brain (Haqq et al., 2003; Strand et al., 2007) and a study has reported that deletion of *ATRNL1* results in cognitive impairment, autism and dysmorphic features (Stark et al., 2010). More importantly, another study has reported that *ATRNL1* is associated with information processing speed and could play a major role in cognitive functions (Luciano et al., 2011). Apart from these evidence, it is not clear from the literature survey whether genetic variants in *ATRNL1* have any role in odour perception.

In addition to the borderline significant associations on chromosome 10, a SNP on chromosome 8 (rs12543038, $P = 1.7 \times 10^{-6}$, $OR_{Dom} = 1.88$, Appendix E.4) also attained borderline significance. rs12543038 is an intron variant in *BAALC* which is preferentially expressed in the human brain and have been shown to be overexpressed

in patients with acute leukaemia (Tanner et al., 2001). A SNP (rs4330979, P = 6.9 × 10^{-6}, OR_{Dom} = 2.03) in *SLC35F3* also showed suggestive significance with odour perception of banana. Thiamine deficiency has been reported to result in cell death in the olfactory bulb in mice (Hamada et al., 2013) and *SLC35F3* has been shown to be involved in thiamine uptake.

Table 6.5 | Association of genetic variants with banana odour perception in the PoBI data.

SNP	***Chr***	***Position***	***Gene[a]***	***p-value[b]***	***OR***	***r^2***
rs1336115*	10	117328426	*ATRNL1*	3.6×10^{-7}	1.72	-
rs7895853	10	117282016	*ATRNL1*	5.8×10^{-7}	1.65	.97
rs1932582	10	117327420	*ATRNL1*	7.3×10^{-7}	1.68	.99
rs12543038	8	104244309	*BAALC*	1.8×10^{-6}	1.88	-
rs4330979	1	234098168	*SLC35F3*	6.9×10^{-6}	2.03	-

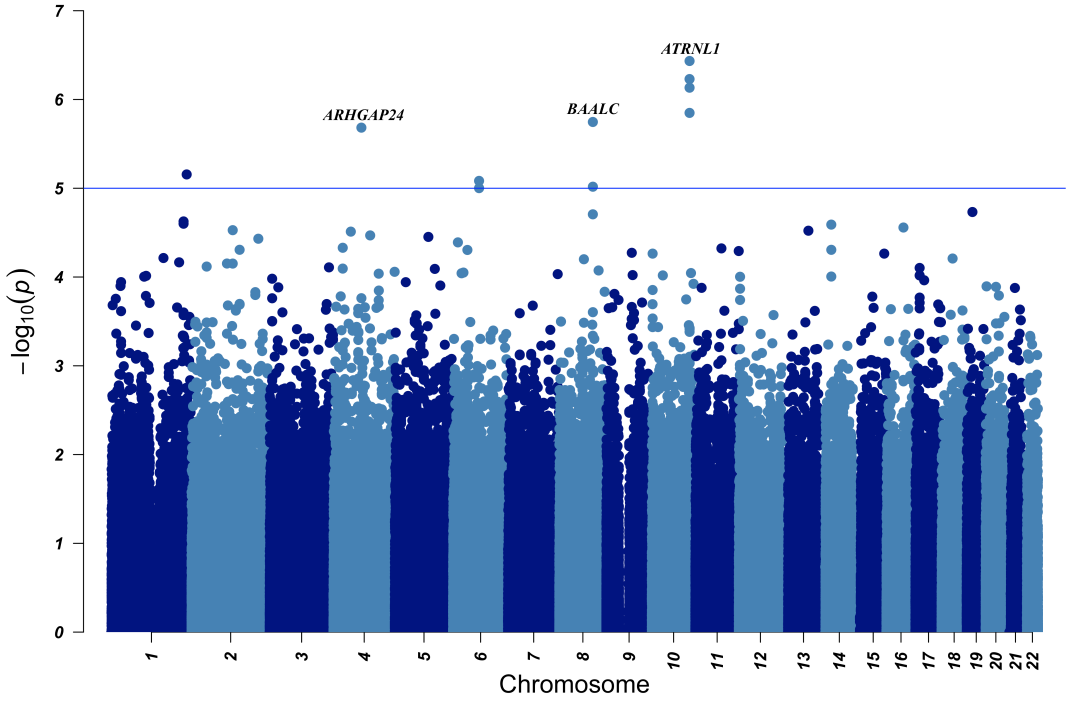
SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; r^2, r-squared *r^2 calculated between rs1336115 and other SNPs on chromosome 10 with suggestive genome-wide significance); [b]Fisher's exact test; OR, Odds ratio

Table 6.6 | Distribution of rs1336115 genotypes in participants of banana odour perception.

Genotype	***AA***	***Aa***	***aa***	***Total***
Not smelling	20	170	153	343
Smelling	87	321	566	974
Total	107	491	719	1317

$P_{Dom} = 1.7\times10^{-5}$, $OR_{Dom} = 1.72$, $P_{Rec} = 0.08415$, $OR_{Rec} = 1.58$

Figure 6.8 | Manhattan plot for the GWAS of banana odour perception. The observed $-log_{10}$ transformed -values (Fisher's exact test) are plotted (y axis) against their chromosomal positions (x axis). Each dot represents a SNP on their chromosomal location. The horizontal blue line indicates the threshold for suggestive genome-wide significance ($P = 1 \times 10^{-5}$)



6.6 **Discussion**

Olfactory receptors are the largest family of receptors in humans comprising ~ 400 olfactory genes and ~600 olfactory pseudo genes. It is important to note that number of olfactory receptors vary species-to-species as they are significantly enriched in regions of CNVs (Hasin-Brumshtein et al., 2009). The proteins encoded by these receptors, or their combinations enable us to distinguish between different odours related to food and other substances. The olfactory system in humans is thought to influence food selection, social behaviour and could play a major role in day-to-day life. Yet the genetics underlying olfaction is poorly understood. The present study identified four genetic variants that showed suggestive genome-wide associations supported by strong biological evidences with odour perception.

6.6.1 Protocadherins, autism and sense of smell

The SNP rs10507739 in *PCDH9* (a protocadherin) showed suggestive association with odour perception of chocolate. Protocadherins, trans-membrane proteins of the cadherin protein family, are calcium-dependent adhesion molecules that play a fundamental role in cell-cell interaction, synaptogenesis and cell signalling during embryonic development (Asahina et al., 2012). *PCDH9*, highly expressed in the nervous system and olfactory bulb (OB), has been shown to be associated with several neurological disorders, including autism spectrum disorder, where cognitive impairment is a common feature (Marshall et al., 2008). *PCDH9* has also been shown to be associated (suggestive genome-wide significance) with the sense of smell in a genome-wide meta-analysis (Dong et al., 2015), which suggests that it could be a genetic component in odour perception. It is important to note that poor sense of smell

is an early sign of neurological disorders and patients with autism spectrum disorder have been shown to have increased olfactory dysfunction (Rozenkrantz et al., 2015). Notably, both olfactory neurons and protocadherins work in a calcium-dependent manner (Menini, 1999). Taken together, all this evidence suggests that *PCDH9* may be a genetic component behind the sense of smell. Other than the suggestive association shown in a meta-analysis (Dong et al., 2015), it is the first time that *PCDH9* has been reported to have suggestive association with the sense of smell.

6.6.2 Chromosome 15q11-13, olfactory receptors and neurological disorders

A SNP (rs11632280,) within 300 kb of *PWRN4* and *NDN* in the current analysis showed suggestive genome-wide significance with odour perception of chocolate. rs11632280 lies within the chromosome 15q11-13 region that is known for cytogenetic abnormalities such as deletion, duplication, increased recombination frequencies and genomic imprinting (Shao et al., 2003). These chromosomal abnormalities have been shown to be associated with different neurological disorders, including Angelman syndrome (AS), Prader-Willi syndrome (PWS), autism disorder and ataxia (Bass et al., 2000; Buiting, 2010). Individuals with Prader-Willi syndrome (PWS) tend to show hyperphagia and reduced selectivity in choosing food items for consumption. Interestingly, studies have shown that taste and olfaction may influence food preferences (Eriksson et al., 2012). Furthermore, a cluster of olfactory receptors lies on the same chromosomal arm (15q11-13) which may explain the lessened selectivity in eating behaviour of Prader-Willi syndrome cases (Bittel & Butler, 2005).

Additionally, a cluster of GABA receptors (*GABRB3*, *GABRA5*, *GABRG3*) is located ~3 Mb from *NDN* and *PWRN4*, although the SNPs in *GABRB3* are not in LD with the most significant hit (rs11632280, $r^2 =$ 0.00020). These genes encode receptors for gamma-aminobutyric acid (GABA, a major inhibitory neurotransmitter in the mammalian central nervous system). The periglomerular (PG) cells in the olfactory bulb serve as a bridge between glomeruli and mitral cells to process detected odours to the olfactory cortex. It has been shown that all periglomerular (PG) cells in the rodent olfactory bulb express GABA receptors (Panzanelli et al., 2007); therefore, a potential role of these receptors in odour perception may exist. In summary, rs11632280 could be a genetic component of odour perception as: 1) rs11632280 lies in the Prader-Willi region on chromosome 15 and reduced selectivity in choosing food items in Prader-Willi syndrome cases could be due to the poor sense of taste and smell as olfactory receptors *OR4M2* & *OR4N4* lie in the proximity of this region and 2) a consistent olfactory impairment has been shown to be associated with neurological disorders and chromosome 15q11-13 region is cytogenetically abnormal in autism and PWS.

rs227722 on chromosome 17 showed suggestive genome-wide significance with smoke odour perception. A gene within this region (~200 kb) codes for a secreted factor called Noggin that regulates bone morphogenetic protein (BMP) in the mammalian brain (Bond et al., 2012). Noggin protein regulates adult neurogenesis and inhibits the BMP activity where stem cells differentiate into neuroblasts in the olfactory bulb (Lim et al., 2000). Most importantly, rs227722 is in the proximity of olfactory receptors on chromosome 17 (q22). Although rs227722 and the olfactory receptors on chromosome 17 are quite distant and are not in LD ($r^2 = 0.00014$), a

potential cis-interaction between these two could underlie its association with odour perception.

For the perception of odour in banana, a SNP (rs1336115, $P = 3.7 \times 10^{-7}$, $OR_{Dom} = 1.72$) on chromosome 10 in *ATRNL1* showed suggestive genome-wide significance. *ATRNL1* (Attractin-like protein 1) is a single-pass membrane protein which has been shown to be associated with information processing speed and could play a major role in cognitive functions.

6.7 **Conclusions**

Despite the large number of mammalian olfactory receptors, odour perception is poorly understood. The present study provides novel suggestions that *PCDH9*, Necdin protein, Noggin protein and *ATRNL1* can be the genetic components underlying sense of smell, and further studies with large sample size are needed to validate these findings. The genetic variants suggestively associated with sense of smell identified in this study have also been shown to be associated with different neurological disorders, including PWS, AS and autism spectrum disorder. It is important to note that olfactory impairment is an onset of many neurological disorders and, as shown in the current data, a possible role of olfactory components in these disorders cannot be denied.

7 Discussion

My research began with the aim to investigate the genetic basis of human pigmentation, taste and olfactory perception as well as to identify genetic variants underlying individuals' food preferences, in a well-characterised population of the British Isles. The fundamental focus was to search for the novel genetic variants as well as to replicate previously known genetic factors underlying normal human features such as hair and skin pigmentation. The phenotypes chosen in the present research cover a wide range of normal human traits such as hair and skin pigmentation, sensory perception and food preferences, and the novel findings of my research will add new insights to fields including pigmentation research, sensory science and nutri-genetics.

A total of 1,593 unrelated samples (which had phenotype and genotype data available) from the British Isles were used to study the genetic basis of normal human traits which are broadly divided into three categories: 1) pigmentation; 2) sensory perception (taste and olfaction) and 3) food preferences. In order to investigate that there are genetic variants conferring large effects on the studied phenotypes in the current analysis, the phenotypes were dichotomised. For example, based on the Figure 3.3, lower extremes (~15% of total individuals) were tested for genetic associations. The results were consistent with the previously known genetic variants for hair and skin pigmentation as well as with taste perception, and did not find any genome-wide significant variants associated with food preference data. However, for food preference data, all of the associations (SNPs that attained suggestive genome-wide significance, $P = 1 \times 10^{-5}$) were supported by biological evidence available in the literature, and thus the fact that they attain p-values of the order of 10^{-5} constitutes

statistical support for their involvement, even though this is not considered 'genome wide significant' according to standard GWAS assumptions.

For normal hair pigmentation, this research replicated nearly all of the well documented pigmentation genes, including *MC1R*, *SLC45A2*, *TYR*, *IRF4*, *HERC2*, *KITLG* and *SLC24A4* and results generally agreed with the studies performed in the past (Adhikari et al., 2016; Dotson et al., 2008; Eriksson et al., 2010b; Jacobs et al., 2015; Sulem et al., 2007). The present study did not report any novel genome-wide significant genetic associations of hair and skin colour. However, genetic variants in *KLF15* (red hair), *KL* (blond hair) and *NLN* (hair greying) attained suggestive genome-wide significance with relatively large odds ratios, which could suggest their potential involvement in determining the studied trait.

Pigmentation is considered to be a complex trait, but even well documented genes, for example *OCA2* / *HERC2*, *IRF4*, *SLC24A4*, *SLC45A2*, and *MC1R* only explain ~28% phenotypic variance of the hair colour (blond to brown) within northern Europe (Eriksson et al., 2010b) suggesting that other factors underlie the remaining phenotypic variance. My research suggests a possible role of variants in *KLF15* and an intergenic variant near *KL* (<200 kb, rs17643831) in hair pigmentation (OR range 3.0 - 4.9). Importantly, the C allele of rs17643831 (near *KL* region) is common in Europeans and rare in Africans (completely absent in Yoruba (YRI) from HapMap). In addition, klotho (*KL*) has been shown to regulate pigment synthesis by regulating MITF and TYR in cultured human retinal pigment epithelium (Kokkinaki, 2016). Together with the large effect sizes, it therefore seems likely that *KLF15* and especially *KL* may have a key role in pigmentation biology.

Hair greying is an indication of ageing in humans, but the underlying genetics behind

variation in this phenotype is not clear. A GWAS performed in admixed Latin Americans reported *IRF4* (rs12203592) as a determinant of hair greying in humans (Adhikari et al., 2016). rs12203592 replicated ($P < 0.05$), but did not show genome-wide significance in the current analysis (rs12203592, $P = 4 \times 10^{-5}$), and the most significant SNP was an intron variant in *NLN* (neurolysin, $P = 1.5 \times 10^{-7}$, $OR_{Rec} = 17.08$) on chromosome 5. The fact that rs12203592 showed genome-wide significance in the previous study (Adhikari et al., 2016), but not in the PoBI data, could be due to the ethnic differences in the populations analysed or the way the phenotype data were collected. Previous studies have shown that the age of hair greying differs according to ethnic/geographical origin, and the hair greying data collected in the current analysis are roughly consistent with the 50/50/50 rule of thumb which states that 'at age 50 years, 50% of the population has at least 50% grey hair' (Keogh & Walsh, 1965; Panhard et al., 2012) (Figure 3.3). Nevertheless, the results for rs12203592 ($P = 4 \times 10^{-5}$) constitute a valid replication. Even accounting for the fact that there are in the region of 100 candidate SNPs, a Bonferroni correction gives ($P = (4 \times 10^{-5}) \times 100 = 0.004$) which is significant at the 5% level.

Although the most significant SNP for hair greying did not attain genome-wide significance, literature suggests that *NLN* could be a potential determinant of hair greying. Neuorlysin is a zinc metallopeptidase which is known to be involved in neuropeptide and especially neurotensin metabolism (Brown et al., 2001). Interestingly, it has been suggested that neurotensin can modulate secretion of the pituitary hormone (for example, alpha-MSH) through regulation of the hypothalamic neurons (Desrues et al., 1998).

Using objective measurements (with a skin colormeter) of skin melanin and erythema

levels, this research confirms association of the melanin synthesis gene *SLC45A2* with skin pigmentation, replicating previous research performed using the same approach (Stokowski et al., 2007). In addition, the strong biological evidence available in the literature and the relatively large effect sizes (OR range 1.8 - 2.4), suggest that borderline significant associations of *CHEK1* (associated with skin melanin levels), *PTPRQ* and *SLC27A3* (associated with skin erythema levels) may underlie visible colour differences among human populations. Interestingly, for rs4341393 (P = 3.1×10^{-7}, OR_{Dom} = 2.44, Appendix B.16) which lies in *GATAD2B*, the allele that has a positive association with skin redness (T), is common in Europeans (51%) and relatively rare in Asians (7%).

For perception of the bitter tasting compounds PTC and PROP, a strong peak of SNPs in the *TAS2R38* region showed genome-wide signals and these results are consistent with the literature (Reed et al., 2010). In contrast to PTC and PROP, people can perceive sodium benzoate (commonly used as food preservative) as sweet (in the current dataset 66.22% participant tasted as sweet, n (total) = 1,513), sour, salty or bitter. For sodium benzoate taste perception, two variants near (<1.5 Mb) taste receptor genes *TAS1R3* and *TAS1R1,* showed relatively large odds ratios (OR range 1.9 – 4.1). Importantly, it has been shown that preference for saccharin in mice is mediated by *TAS1R3* (Reed et al., 2004), although the variants in *TAS1R3* and *TAS1R1* showed effectively no LD with the top hit for sodium benzoate taste perception. This is the first GWAS report of sodium benzoate taste perception.

There has been a growing interest in understanding the genetic factors that underlie human dietary intake as our food preferences may result in nutrition-related chronic disease or obesity (Smith et al., 2016). Preference ratings for seven foods, that

represent a variety of tastes, were also analysed in this thesis. There were no genome-wide associations with food preferences, which may implicate that they are largely determined by environmental factors rather than genetics or a combination of both. However, there were suggestive associations, including *MAGI2, PTPRD, KIF3C* and *KCNK1* (OR range 1.8 -10.89) for influencing certain food preferences. A variant in *MAGI2* showed strong evidence for Marmite preference (OR = 10.89) and this association could implicate a possible biological role of MAGI2 in taste perception as this brain-specific scaffolding protein interacts with glutamate receptors (NMDA receptors) (Deng et al., 2006a). Here, it is important to note that monosodium glutamate (sodium salt of glutamic acid, which is contained in the Marmite recipe) is a well-known flavour-enhancing compound and sensitivity to MSG shows bimodal distribution in humans (Jobling, 2014a) as is the case in the PoBI data. Importantly, the previous findings that NMDA serves as a primary taste transducer for MSG (Dingledine & Conn, 2000) and that *MAGI2* plays a key role in synaptic trafficking of NMDA and AMPA receptors (Deng et al., 2006b), the current findings suggest that polymorphism in *MAGI2* may mediate Marmite preference.

This study also showed novel suggestive associations with odour perception, including SNPs within or nearby (<0.80 Mb) *PCDH9*, *PWRN4*/*NDN*, *CDH11*, *NFIB* and *ATRNL1* (Odds ratio range 1.7 – 4.1). Olfactory perception has a major role in day-to-day life, but this field is still poorly understood and only a handful of studies have been performed to investigate its genetic basis. Although a genome-wide meta-analysis has shown that *PCDH9* is suggestively associated with sense of smell, this is the first GWAS study reporting a potential role of *PWRN4*/*NDN*, *CDH11*, *NFIB* and *ATRNL1* in odour perception based on genome-wide suggestive signals (Dong et al., 2015). *ATRNL1, PCDH9* and *NFIB* are expressed highly in the human brain. All these

association signals have biological implications in cells related to the olfactory system and thus makes them good candidates for future studies. *PCDH9, PWRN4* and *ATRNL1* have been shown to be involved in different neurological disorders such as autism spectrum disorder, Parkinson's and Alzheimer's disease and studies have also demonstrated that olfactory dysfunction is a common mechanism in neurological disorders (Rozenkrantz et al., 2015). In addition, Prader–Willi syndrome (PWS) patients show reduced selectivity in choosing food items for consumption, and metabolic disorders such as obesity and type-2 diabetes are quite common among them. Most importantly, a cluster of olfactory receptors lies on chromosome 15(q11), ~2 Mb of *PWRN4,* and deletion of 15 (q11-13) is associated with PWS and Angelman syndrome (AS). Taken together, this evidence raises the question of whether the reduced selectivity in choosing food items for consumption in PWS patients is due to olfactory impairment. It would be interesting to see how taste and olfactory preferences affect an individual's health and whether there are any potential links between chemosensory perception and different metabolic and neurological disorders such as type 2 diabetes and Parkinson's disease.

Overall, my research replicates all of the previously-known genetic variants in hair and skin pigmentation as well as in bitter taste perception, and provides suggestive variants which could be the potential determinants of pigmentation, taste, olfaction and food preferences. It should be taken into account that these results, although supported by biological evidence and relatively large odds ratios, are preliminary and further studies would be required to validate. Genetics of food preferences is an undeveloped field, albeit a handful of studies (Pallister et al., 2015; Pirastu et al., 2016), and therefore it is difficult to draw strong conclusions based on these results.

Although I believe that the findings presented have enough evidence to support their suggested biological roles, this research has a number of possible limitations: 1) As discussed in the chapter 1, replication of GWAS findings is important to confirm that observed genetic associations underlie the phenotype of interest. Further research is required to do so for the suggestively associated findings that have been presented, although several previous findings (mostly for hair colour and PTC/PROP taste perception) were successfully replicated. A large number of PoBI samples with phenotype data available have not yet been genotyped and these data can be used to confirm the findings presented here. Additionally, the TwinsUK registry has food preference data which would give an opportunity to perform a replication study

2) The phenotype data collection, for example hair greying, could have been done using a point-scale rather than asking the subjects to remember the age they first noticed hair greying (although the current data are, roughly, consistent with the previous research and as shown in the Figure 3.3 follows the 50/50/50 rule of thumb in context to hair greying), as many individuals struggled to remember a specific age, or were confused about what stage of the greying process was being enquired about.

3) The samples were genotyped on two different chips (Illumina Human 1.2M-Duo and Infinium OmniExpress-24 bead array) which is not ideal for a GWAS design (although implementing stringent QC filters could overcome this problem)

4) Genotype imputation would in theory have been useful for detecting the causal variants observed in the various analyses, but it could not be done due to time constraints.

In summary, I have taken advantage of a sampling criterion where individuals were targeted on the basis of having four grandparents from the same rural area, minimizing the risk of spurious genotype-phenotype correlations that are due to population stratification. Emphasis on careful collection and selection of the appropriate phenotypic categories has led to results that are both statistically and biologically persuasive, with large genetic effects.

8 References

Abbott, M. A., Wells, D. G., & Fallon, J. R. (1999). The insulin receptor tyrosine kinase substrate p58/53 and the insulin receptor are components of CNS synapses. *Journal of Neuroscience, 19*(17), 7300-7308.

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacon-Duque, J. C., . . . Ruiz-Linares, A. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications, 7*. doi:ARTN 10815 10.1038/ncomms10815

Akins, M. R., Benson, D. L., & Greer, C. A. (2007). Cadherin expression in the developing mouse olfactory system. *Journal of Comparative Neurology, 501*(4), 483-497. doi:10.1002/cne.21270

Allentoft, M. E., Sikora, M., Sjogren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., . . . Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. *Nature, 522*(7555), 167-172. doi:10.1038/nature14507

Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. (2001). Genomewide scans of complex human diseases: True linkage is hard to find. *American Journal of Human Genetics, 69*(5), 936-950. doi:Doi 10.1086/324069

Altshuler, D., Brooks, L. D., Chakravarti, A., Collins, F. S., Daly, M. J., Donnelly, P., . . . Consortium, I. H. (2005). A haplotype map of the human genome. *Nature, 437*(7063), 1299-1320. doi:10.1038/nature04226

Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., . . . Consortium, G. P. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-+. doi:10.1038/nature15393

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., & Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nat Protoc, 5*(9), 1564-1573. doi:10.1038/nprot.2010.116

Ando, H., Ryu, A., Hashimoto, A., Oka, M., & Ichihashi, M. (1998). Linoleic acid and alpha-linolenic acid lightens ultraviolet-induced hyperpigmentation of the skin. *Arch Dermatol Res, 290*(7), 375-381.

Andrieu, D., Meziane, H., Marly, F., Angelats, C., Fernandez, P. A., & Muscatelli, F. (2006). Sensory defects in Necdin deficient mice result from a loss of sensory neurons correlated within an increase of developmental programmed cell death. *Bmc Developmental Biology, 6*. doi:Artn 56 10.1186/1471-213x-6-56

Antonopoulou, K., Stefanaki, I., Lill, C. M., Chatzinasiou, F., Kypreou, K. P., Karagianni, F., . . . Stratigos, A. J. (2015). Updated Field Synopsis and Systematic Meta-Analyses of Genetic Association Studies in Cutaneous Melanoma: The MelGene Database. *Journal of Investigative Dermatology, 135*(4), 1074-1079. doi:10.1038/jid.2014.491

Asahina, H., Masuba, A., Hirano, S., & Yuri, K. (2012). Distribution of protocadherin 9 protein in the developing mouse nervous system. *Neuroscience, 225*, 88-104. doi:10.1016/j.neuroscience.2012.09.006

Asif, M. (2014). The prevention and control the type-2 diabetes by changing lifestyle and dietary pattern. *J Educ Health Promot, 3*, 1. doi:10.4103/2277-9531.127541

Astle, W., & Balding, D. J. (2009). Population Structure and Cryptic Relatedness in Genetic Association Studies. *Statistical Science, 24*(4), 451-471. doi:10.1214/09-Sts307

Barsh, G. S. (2003). What controls variation in human skin color? *PLoS Biol, 1*(1), E27. doi:10.1371/journal.pbio.0000027

Bartel, D. L., Sullivan, S. L., Lavoie, E. G., Sevigny, J., & Finger, T. E. (2006). Nucleoside triphosphate diphosphohydrolase-2 is the ecto-ATPase of type I cells in taste buds. *Journal of Comparative Neurology, 497*(1), 1-12. doi:10.1002/cne.20954

Bashir, Z. I., Alford, S., Davies, S. N., Randall, A. D., & Collingridge, G. L. (1991). Long-term potentiation of NMDA receptor-mediated synaptic transmission in the hippocampus. *Nature, 349*(6305), 156-158. doi:10.1038/349156a0

Bass, M. P., Menold, M. R., Wolpert, C. M., Donnelly, S. L., Ravan, S. A., Hauser, E. R., . . . Pericak-Vance, M. A. (2000). Genetic studies in autistic disorder and chromosome 15. *Neurogenetics, 2*(4), 219-226. doi:DOI 10.1007/s100480050067

Beleza, S., Johnson, N. A., Candille, S. I., Absher, D. M., Coram, M. A., Lopes, J., . . . Tang, H. (2013). Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet, 9*(3), e1003372. doi:10.1371/journal.pgen.1003372

Bellono, N. W., Escobar, I. E., & Oancea, E. (2016). A melanosomal two-pore sodium channel regulates pigmentation. *Scientific Reports, 6*. doi:ARTN 26570
10.1038/srep26570

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., . . . Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*(7218), 53-59. doi:10.1038/nature07517

Bigdeli, T. B., Neale, B. M., & Neale, M. C. (2014). Statistical properties of single-marker tests for rare variants. *Twin Res Hum Genet, 17*(3), 143-150. doi:10.1017/thg.2014.17

Bishop, D. T., Demenais, F., Iles, M. M., Harland, M., Taylor, J. C., Corda, E., . . . Bishop, J. A. N. (2009). Genome-wide association study identifies three loci associated with melanoma risk. *Nature Genetics, 41*(8), 920-U985. doi:10.1038/ng.411

Bittel, D. C., & Butler, M. G. (2005). Prader-Willi syndrome: clinical genetics, cytogenetics and molecular biology. *Expert Rev Mol Med, 7*(14), 1-20. doi:10.1017/S1462399405009531

Bodmer, W., & Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet, 40*(6), 695-701. doi:10.1038/ng.f.136

Bodmer, W. F. (1973). Genetic factors in Hodgkin's disease: association with a disease-susceptibility locus (DSA) in the HL-A region. *Natl Cancer Inst Monogr, 36*, 127-134.

Bond, A. M., Bhalala, O. G., & Kessler, J. A. (2012). The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. *Dev Neurobiol, 72*(7), 1068-1084. doi:10.1002/dneu.22022

Branicki, W., Liu, F., van Duijn, K., Draus-Barini, J., Pospiech, E., Walsh, S., . . . Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. *Hum Genet, 129*(4), 443-454. doi:10.1007/s00439-010-0939-8

Brauer, G., & Chopra, V. P. (1978). [Estimation of the heritability of hair and eye color]. *Anthropol Anz, 36*(2), 109-120.

Breslin, P. A., & Spector, A. C. (2008). Mammalian taste perception. *Curr Biol, 18*(4), R148-155. doi:10.1016/j.cub.2007.12.017

Brookes, A. J. (1999). The essence of SNPs. *Gene, 234*(2), 177-186. doi:Doi 10.1016/S0378-1119(99)00219-X

Brown, C. K., Madauss, K., Lian, W., Beck, M. R., Tolbert, W. D., & Rodgers, D. W. (2001). Structure of neurolysin reveals a deep channel that limits substrate access. *Proceedings of the National Academy of Sciences of the United States of America, 98*(6), 3127-3132. doi:DOI 10.1073/pnas.051633198

Brunet, L. J., Gold, G. H., & Ngai, J. (1996). General anosmia caused by a targeted disruption of the mouse olfactory cyclic nucleotide-gated cation channel. *Neuron, 17*(4), 681-693. doi:Doi 10.1016/S0896-6273(00)80200-7

Buck, L., & Axel, R. (1991). A Novel Multigene Family May Encode Odorant Receptors - a Molecular-Basis for Odor Recognition. *Cell, 65*(1), 175-187. doi:Doi 10.1016/0092-8674(91)90418-X

Buiting, K. (2010). Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet C Semin Med Genet, 154C*(3), 365-376. doi:10.1002/ajmg.c.30273

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., . . . Collaborat, B. C. S. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*(7145), 661-678. doi:10.1038/nature05911

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *Plos Computational Biology, 8*(12). doi:ARTN e1002822 10.1371/journal.pcbi.1002822

Callaway, E. (2012). Evolutionary biology: the lost appetites. *Nature, 486*(7403), S16-17. doi:10.1038/486S16a

Campa, D., De Rango, F., Carrai, M., Crocco, P., Montesanto, A., Canzian, F., . . . Barale, R. (2012). Bitter Taste Receptor Polymorphisms and Human Aging. *Plos One, 7*(11). doi:ARTN e45232 10.1371/journal.pone.0045232

Candille, S. I., Absher, D. M., Beleza, S., Bauchet, M., McEvoy, B., Garrison, N. A., . . . Shriver, M. D. (2012). Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations. *PLoS One, 7*(10), e48294. doi:10.1371/journal.pone.0048294

Chandrashekar, J., Hoon, M. A., Ryba, N. J. P., & Zuker, C. S. (2006). The receptors and cells for mammalian taste. *Nature, 444*(7117), 288-294. doi:10.1038/nature05401

Chandrashekar, J., Kuhn, C., Oka, Y., Yarmolinsky, D. A., Hummler, E., Ryba, N. J., & Zuker, C. S. (2010). The cells and peripheral representation of sodium taste in mice. *Nature, 464*(7286), 297-301. doi:10.1038/nature08783

Chandrashekar, J., Mueller, K. L., Hoon, M. A., Adler, E., Feng, L. X., Guo, W., . . . Ryba, N. J. P. (2000). T2Rs function as bitter taste receptors. *Cell, 100*(6), 703-711. doi:Doi 10.1016/S0092-8674(00)80706-0

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4*, 7. doi:10.1186/s13742-015-0047-8

Chang, C. Q., Yesupriya, A., Rowell, J. L., Pimentel, C. B., Clyne, M., Gwinn, M., . . . Schully, S. D. (2014). A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. *Eur J Hum Genet, 22*(3), 402-408. doi:10.1038/ejhg.2013.161

Chaudhari, N., Landin, M. A., & Roper, S. D. (2000). A metabotropic glutamate receptor variant functions as a taste receptor. *Nature Neuroscience, 3*(2), 113-119. doi:Doi 10.1038/72053

Chaudhari, N., & Roper, S. D. (2010). The cell biology of taste. *J Cell Biol, 190*(3), 285-296. doi:10.1083/jcb.201003144

Chaudhari, N., Yang, H., Lamp, C., Delay, E., Cartford, C., Than, T., & Roper, S. (1996). The taste of monosodium glutamate: Membrane receptors in taste buds. *Journal of Neuroscience, 16*(12), 3817-3826.

Chen, X., & Sullivan, P. F. (2003). Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J, 3*(2), 77-96. doi:10.1038/sj.tpj.6500167

Chen, Y. T., Lin, W. D., Liao, W. L., Lin, Y. J., Chang, J. G., & Tsai, F. J. (2015). PTPRD silencing by DNA hypermethylation decreases insulin receptor signaling and leads to type 2 diabetes. *Oncotarget, 6*(15), 12997-13005.

Chow, C. Y., Zhang, Y. L., Dowling, J. J., Jin, N., Adamska, M., Shiga, K., . . . Meisler, M. H. (2007). Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. *Nature, 448*(7149), 68-72. doi:10.1038/nature05876

Clarke, G. M., Anderson, C. A., Pettersson, F. H., Cardon, L. R., Morris, A. P., & Zondervan, K. T. (2011). Basic statistical analysis in genetic case-control studies. *Nat Protoc, 6*(2), 121-133. doi:10.1038/nprot.2010.182

Clayton, D. G. (2009). Sex chromosomes and genetic association studies. *Genome Med, 1*(11), 110. doi:10.1186/gm110

Comuzzie, A. G., Cole, S. A., Laston, S. L., Voruganti, V. S., Haack, K., Gibbs, R. A., & Butte, N. F. (2012). Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. *Plos One, 7*(12). doi:ARTN e51954
10.1371/journal.pone.0051954

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., . . . Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature, 464*(7289), 704-712. doi:10.1038/nature08516

Cookson, W., Liang, L., Abecasis, G., Moffatt, M., & Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nat Rev Genet, 10*(3), 184-194. doi:10.1038/nrg2537

Cooper, G. M., Johnson, J. A., Langaee, T. Y., Feng, H., Stanaway, I. B., Schwarz, U. I., . . . Rieder, M. J. (2008). A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood, 112*(4), 1022-1027. doi:10.1182/blood-2008-01-134247

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., . . . Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science, 261*(5123), 921-923.

Cox, D. G., & Kraft, P. (2006). Quantification of the power of Hardy-Weinberg equilibrium testing to detect genotyping error. *Hum Hered, 61*(1), 10-14. doi:10.1159/000091787

Crawford, N. G., Kelly, D. E., Hansen, M. E. B., Beltrame, M. H., Fan, S., Bowman, S. L., . . . Tishkoff, S. (2017). Loci associated with skin pigmentation identified in African populations. *Science*. doi:10.1126/science.aan8433

Darwin, C. (1871). The Descent of Man

Davenport, G. C., & Davenport, C. B. (1907). Heredity of Eye-Color in Man. *Science, 26*(670), 589-592. doi:10.1126/science.26.670.589-b

De Angelis, D. A. (1999). Why FRET over genomics? *Physiol Genomics, 1*(2), 93-99.

De Caterina, R., Zampolli, A., Del Turco, S., Madonna, R., & Massaro, M. (2006). Nutritional mechanisms that influence cardiovascular disease. *Am J Clin Nutr, 83*(2), 421S-426S.

DeMaria, S., & Ngai, J. (2010). The cell biology of smell. *Journal of Cell Biology, 191*(3), 443-452. doi:10.1083/jcb.201008163

Deng, F., Price, M. G., Davis, C. F., Mori, M., & Burgess, D. L. (2006a). Stargazin and other transmembrane AMPA receptor regulating proteins interact with synaptic scaffolding protein MAGI-2 in brain. *J Neurosci, 26*(30), 7875-7884. doi:10.1523/JNEUROSCI.1851-06.2006

Deng, F., Price, M. G., Davis, C. F., Mori, M., & Burgess, D. L. (2006b). Stargazin and other transmembrane AMPA receptor regulating proteins interact with synaptic scaffolding protein MAGI-2 in brain. *Journal of Neuroscience, 26*(30), 7875-7884. doi:10.1523/Jneurosci.1851-06.2006

Desrues, L., Tonon, M. C., Leprince, J., Vaudry, H., & Conlon, J. M. (1998). Isolation, primary structure, and effects on alpha-melanocyte-stimulating hormone release of frog neurotensin. *Endocrinology, 139*(10), 4140-4146. doi:10.1210/endo.139.10.6255

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *American Journal of Human Genetics, 65*(4), A83-A83.

Dingledine, R., & Conn, P. J. (2000). Peripheral glutamate receptors: Molecular biology and role in taste sensation. *Journal of Nutrition, 130*(4), 1039s-1042s.

Distefano, J. K., & Taverna, D. M. (2011). Technological issues and experimental design of gene association studies. *Methods Mol Biol, 700*, 3-16. doi:10.1007/978-1-61737-954-3_1

Dong, J., Yang, J. Y., Tranah, G., Franceschini, N., Parimi, N., Alkorta-Aranburu, G., . . . Chen, H. L. (2015). Genome-wide Meta-analysis on the Sense of Smell Among US Older Adults. *Medicine, 94*(47). doi:ARTN e1892 10.1097/MD.0000000000001892

Dotson, C. D., Zhang, L., Xu, H., Shin, Y. K., Vigues, S., Ott, S. H., . . . Munger, S. D. (2008). Bitter Taste Receptors Influence Glucose Homeostasis. *Plos One, 3*(12). doi:ARTN e3974 10.1371/journal.pone.0003974

Drewnowski, A. (2001). The science and complexity of bitter taste. *Nutr Rev, 59*(6), 163-169.

Eiberg, H., Troelsen, J., Nielsen, M., Mikkelsen, A., Mengel-From, J., Kjaer, K. W., & Hansen, L. (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Human Genetics, 123*(2), 177-187. doi:10.1007/s00439-007-0460-x

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., . . . Mountain, J. (2010a). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet, 6*(6), e1000993. doi:10.1371/journal.pgen.1000993

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., . . . Mountain, J. (2010b). Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. *Plos Genetics, 6*(6). doi:ARTN e1000993
10.1371/journal.pgen.1000993

Eriksson, N., Wu, S., Do, C. B., Kiefer, A. K., Tung, J. Y., Mountain, J. L., . . . Francke, U. (2012). A genetic variant near olfactory receptor genes influences cilantro preference. *Flavour, 1*(1), 22. doi:10.1186/2044-7248-1-22

Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet, 24*(8), 1202-1205. doi:10.1038/ejhg.2015.269

Fellermann, K., Stange, D. E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C. L., . . . Stange, E. F. (2006). A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet, 79*(3), 439-448. doi:10.1086/505915

Fisher, R. A. F., E. B.; HUXLEY, J. (1939). Taste-testing the Anthropoid Apes. *Nature*. doi:10.1038/144750a0

Fox, A. L. (1932). The Relationship between Chemical Constitution and Taste. *Proc Natl Acad Sci U S A, 18*(1), 115-120.

Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., . . . Parkes, M. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet, 42*(12), 1118-1125. doi:10.1038/ng.717

Frazer, K. A., Murray, S. S., Schork, N. J., & Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics, 10*(4), 241-251. doi:10.1038/nrg2554

Furuta, J., Nobeyama, Y., Umebayashi, Y., Otsuka, F., Kikuchi, K., & Ushijima, T. (2006). Silencing of Peroxiredoxin 2 and aberrant methylation of 33 CpG islands in putative promoter regions in human malignant melanomas. *Cancer Research, 66*(12), 6080-6086. doi:10.1158/0008-5472.Can-06-0157

Genetic Analysis of Psoriasis, C., the Wellcome Trust Case Control, C., Strange, A., Capon, F., Spencer, C. C., Knight, J., . . . Trembath, R. C. (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet, 42*(11), 985-990. doi:10.1038/ng.694

Genomes Project, C., Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., . . . McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. *Nature, 467*(7319), 1061-1073. doi:10.1038/nature09534

Gerstenblith, M. R., Shi, J., & Landi, M. T. (2010). Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. *Pigment Cell Melanoma Res, 23*(5), 587-606. doi:10.1111/j.1755-148X.2010.00730.x

Gertrude C. Davenport, C. B. D. (1910). Heredity of Skin Pigmentation in Man. *The American Naturalist*.

Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J., & Hayes, B. J. (2016). Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc Biol Sci, 283*(1835). doi:10.1098/rspb.2016.0569

Guenther, C. A., Tasic, B., Luo, L. Q., Bedell, M. A., & Kingsley, D. M. (2014). A molecular basis for classic blond hair color in Europeans. *Nature Genetics, 46*(7), 748-752. doi:10.1038/ng.2991

Haldane, J. B. S. (1956). The Estimation and Significance of the Logarithm of a Ratio of Frequencies. *Annals of Human Genetics, 20*(4), 309-311.

Hamada, S., Hirashima, H., Imaeda, M., Okamoto, Y., Hamaguchi-Hamada, K., & Kurumata-Shigeto, M. (2013). Thiamine Deficiency Induces Massive Cell Death in the Olfactory Bulbs of Mice. *Journal of Neuropathology and Experimental Neurology, 72*(12), 1193-1202.

Han, J. L., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., . . . Hunter, D. J. (2008). A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *Plos Genetics, 4*(5). doi:ARTN e1000074 10.1371/journal.pgen.1000074

Haqq, A. M., Rene, P., Kishi, T., Khong, K., Lee, C. E., Liu, H., . . . Cone, R. D. (2003). Characterization of a novel binding partner of the melanocortin-4 receptor: attractin-like protein. *Biochem J, 376*(Pt 3), 595-605. doi:10.1042/BJ20031241

Hardman, J. A., Tobin, D. J., Haslam, I. S., Farjo, N., Farjo, B., Al-Nuaimi, Y., . . . Paus, R. (2015). The Peripheral Clock Regulates Human Pigmentation. *Journal of Investigative Dermatology, 135*(4), 1053-1064. doi:10.1038/jid.2014.442

Hasin-Brumshtein, Y., Lancet, D., & Olender, T. (2009). Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet, 25*(4), 178-184. doi:10.1016/j.tig.2009.02.002

Hasselbalch, A. L., Heitmann, B. L., Kyvik, K. O., & Sorensen, T. I. A. (2008). Studies of Twins Indicate That Genetics Influence Dietary Intake. *Journal of Nutrition, 138*(12), 2406-2412. doi:10.3945/jn.108.087668

Hernandez-Pacheco, N., Flores, C., Alonso, S., Eng, C., Mak, A. C., Hunstman, S., . . . Pino-Yanes, M. (2017). Identification of a novel locus associated with skin colour in African-admixed populations. *Sci Rep, 7*, 44548. doi:10.1038/srep44548

Herr, I., & Buchler, M. W. (2010). Dietary constituents of broccoli and other cruciferous vegetables: implications for prevention and therapy of cancer. *Cancer Treat Rev, 36*(5), 377-383. doi:10.1016/j.ctrv.2010.01.002

Hevezi, P., Moyer, B. D., Lu, M., Gao, N., White, E., Echeverri, F., . . . Zlotnik, A. (2009). Genome-Wide Analysis of Gene Expression in Primate Taste Buds Reveals Links to Diverse Processes. *Plos One, 4*(7). doi:ARTN e6395 10.1371/journal.pone.0006395

Hinrichs, A. L., Wang, J. C., Bufe, B., Kwon, J. M., Budde, J., Allen, R., . . . Goate, A. M. (2006). Functional variant in a bitter-taste receptor (hTAS2R16) influences risk of alcohol dependence. *American Journal of Human Genetics, 78*(1), 103-111. doi:Doi 10.1086/499253

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet, 6*(2), 95-108. doi:10.1038/nrg1521

Huang, Y. A., & Roper, S. D. (2010). Intracellular Ca(2+) and TRPM5-mediated membrane depolarization produce ATP secretion from taste receptor cells. *J Physiol, 588*(Pt 13), 2343-2350. doi:10.1113/jphysiol.2010.191106

Ihara, Y., & Aoki, K. (1999). Sexual selection by male choice in monogamous and polygynous human populations. *Theor Popul Biol, 55*(1), 77-93. doi:10.1006/tpbi.1998.1388

Iles, M. M. (2008). What can genome-wide association studies tell us about the genetics of common disease? *Plos Genetics, 4*(2). doi:ARTN e33 10.1371/journal.pgen.0040033

International HapMap, C., Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., . . . Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature, 449*(7164), 851-861. doi:10.1038/nature06258

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution, 39*(1), 57-106. doi:DOI 10.1006/jhev.2000.0403

Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 8962-8968. doi:10.1073/pnas.0914628107

Jackman, A. H., & Doty, R. L. (2005). Utility of a three-item smell identification test in detecting olfactory dysfunction. *Laryngoscope, 115*(12), 2209-2212. doi:10.1097/01.mlg.0000183194.17484.bb

Jackson, I. J. (1997). Homologous pigmentation mutations in human, mouse and other model organisms. *Hum Mol Genet, 6*(10), 1613-1624.

Jacobs, L. C., Hamer, M. A., Gunn, D. A., Deelen, J., Lall, J. S., van Heemst, D., . . . Nijsten, T. (2015). A Genome-Wide Association Study Identifies the Skin Color Genes IRF4, MC1R, ASIP, and BNC2 Influencing Facial Pigmented Spots. *Journal of Investigative Dermatology, 135*(7), 1735-1742. doi:10.1038/jid.2015.62

Jakobsdottir, J., Gorin, M. B., Conley, Y. P., Ferrell, R. E., & Weeks, D. E. (2009). Interpretation of Genetic Association Studies: Markers with Replicated Highly Significant Odds Ratios May Be Poor Classifiers. *Plos Genetics, 5*(2). doi:ARTN e1000337 10.1371/journal.pgen.1000337

Jeon, T. I., Seo, Y. K., & Osborne, T. F. (2011). Gut bitter taste receptor signalling induces ABCB1 through a mechanism involving CCK. *Biochemical Journal, 438*, 33-37. doi:10.1042/Bj20110009

Jeyaraj, D., Haldar, S. M., Wan, X. P., McCauley, M. D., Ripperger, J. A., Hu, K., . . . Jain, M. K. (2012). Circadian rhythms govern cardiac repolarization and arrhythmogenesis. *Nature, 483*(7387), 96-U141. doi:10.1038/nature10852

Jeyaraj, D., Scheer, F. A. J. L., Ripperger, J. A., Halder, S. M., Lu, Y. A., Prosdocimo, D. A., . . . Jain, M. K. (2012). Klf15 Orchestrates Circadian Nitrogen Homeostasis. *Cell Metabolism, 15*(3), 311-323. doi:10.1016/j.cmet.2012.01.020

Jobling, M. A., Hollox, E., Hurles, M., Kivisild, T., & Tyler-Smith, C. . (2014a). *Human Evolutionary Genetics*

Jobling, M. A., Hollox, E., Hurles, M., Kivisild, T., & Tyler-Smith, C. . (2014b). Human Genome Variation *Human Evolutionary Genetics* (pp. 47-48).

Jones L.V., P. D. R., Thurstone L. L. . (1955). Development of a scale for measuring soldiers'food preferences b. *J Food Sci, 20*, 512–520. doi:10.1111/j.1365-2621. 1955.tb16862.x

Kanageswaran, N., Demond, M., Nagel, M., Schreiner, B. S. P., Baumgart, S., Scholz, P., . . . Gisselmann, G. (2015). Deep Sequencing of the Murine Olfactory Receptor Neuron Transcriptome. *Plos One, 10*(1). doi:ARTN e0113170 10.1371/journal.pone.0113170

Kayser, M., Liu, F., Janssens, A. C., Rivadeneira, F., Lao, O., van Duijn, K., . . . van Duijn, C. M. (2008). Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet, 82*(2), 411-423. doi:10.1016/j.ajhg.2007.10.003

Keller, K. L., Steinmann, L., Nurse, R. J., & Tepper, B. J. (2002). Genetic taste sensitivity to 6-n-propylthiouracil influences food preference and reported intake in preschool children. *Appetite, 38*(1), 3-12. doi:10.1006/appe.2001.0441

Keogh, E. V., & Walsh, R. J. (1965). Rate of greying of human hair. *Nature, 207*(999), 877-878.

Khnykin, D., Miner, J. H., & Jahnsen, F. (2011). Role of fatty acid transporters in epidermis: Implications for health and disease. *Dermatoendocrinol, 3*(2), 53-61. doi:10.4161/derm.3.2.14816

Kim, U. K., & Drayna, D. (2005). Genetics of individual differences in bitter taste perception: lessons from the PTC gene. *Clin Genet, 67*(4), 275-280. doi:10.1111/j.1399-0004.2004.00361.x

Kim, U. K., Jorgenson, E., Coon, H., Leppert, M., Risch, N., & Drayna, D. (2003). Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science, 299*(5610), 1221-1225. doi:DOI 10.1126/science.1080190

King, R. A., Townsend, D., Oetting, W., Summers, C. G., Olds, D. P., White, J. G., & Spritz, R. A. (1991). Temperature-sensitive tyrosinase associated with peripheral pigmentation in oculocutaneous albinism. *J Clin Invest, 87*(3), 1046-1053. doi:10.1172/JCI115064

Kinnamon, S. C. (2012). Taste receptor signalling - from tongues to lungs. *Acta Physiol (Oxf), 204*(2), 158-168. doi:10.1111/j.1748-1716.2011.02308.x

Kinnamon, S. C., & Margolskee, R. F. (1996). Mechanisms of taste transduction. *Current Opinion in Neurobiology, 6*(4), 506-513. doi:Doi 10.1016/S0959-4388(96)80057-2

Klein, R. J. (2007). Power analysis for genome-wide association studies. *BMC Genet, 8*, 58. doi:10.1186/1471-2156-8-58

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., . . . Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science, 308*(5720), 385-389. doi:10.1126/science.1109557

Kokkinaki. (2016). Klotho Regulates Retinal Pigment Epithelial Functions and Protects Against Oxidative Stress (vol 33, pg 16346, 2013). *Journal of Neuroscience, 36*(24), 6597-6597. doi:10.1523/Jneurosci.1578-16.2016

Kokkinaki, M., Abu-Asab, M., Gunawardena, N., Ahern, G., Javidnia, M., Young, J., & Golestaneh, N. (2013). Klotho Regulates Retinal Pigment Epithelial Functions and Protects Against Oxidative Stress. *Journal of Neuroscience, 33*(41), 16346-16359. doi:10.1523/Jneurosci.0402-13.2013

Lander, E. S. (1996). The new genomics: global views of biology. *Science, 274*(5287), 536-539.

Lander, E. S., Consortium, I. H. G. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., . . . Conso, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921. doi:10.1038/35057062

Lapis, T. J., Penner, M. H., & Lim, J. (2016). Humans Can Taste Glucose Oligomers Independent of the hT1R2/hT1R3 Sweet Taste Receptor. *Chemical Senses, 41*(9), 755-762. doi:10.1093/chemse/bjw088

Ledda, M., Kutalik, Z., Destito, M. C. S., Souza, M. M., Cirillo, C. A., Zamboni, A., . . . Genick, U. K. (2014). GWAS of human bitter taste perception identifies new loci and reveals additional complexity of bitter taste genetics. *Human Molecular Genetics, 23*(1), 259-267. doi:10.1093/hmg/ddt404

Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., . . . Genetics, I. M. S. (2015). The fine-scale genetic structure of the British population. *Nature, 519*(7543), 309-+. doi:10.1038/nature14230

Lewontin, R. C., & Kojima, K. (1960). The Evolutionary Dynamics of Complex Polymorphisms. *Evolution, 14*(4), 458-472. doi:Doi 10.2307/2405995

Lim, D. A., Tramontin, A. D., Trevejo, J. M., Herrera, D. G., Garcia-Verdugo, J. M., & Alvarez-Buylla, A. (2000). Noggin antagonizes BMP signaling to create a niche for adult neurogenesis. *Neuron, 28*(3), 713-726.

Lin, B. D., Mbarek, H., Willemsen, G., Dolan, C. V., Fedko, I. O., Abdellaoui, A., . . . Hottenga, J. J. (2015). Heritability and Genome-Wide Association Studies for Hair Color in a Dutch Twin Family Based Sample. *Genes (Basel), 6*(3), 559-576. doi:10.3390/genes6030559

Lin, J. Y., & Fisher, D. E. (2007). Melanocyte biology and skin pigmentation. *Nature, 445*(7130), 843-850. doi:10.1038/nature05660

Lin, W. H., & Kinnamon, S. C. (1999). Physiological evidence for ionotropic and metabotropic glutamate receptors in rat taste cells. *Journal of Neurophysiology, 82*(5), 2061-2069.

Liu, F., Visser, M., Duffy, D. L., Hysi, P. G., Jacobs, L. C., Lao, O., . . . Kayser, M. (2015). Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Hum Genet, 134*(8), 823-835. doi:10.1007/s00439-015-1559-0

Liu, F., Wollstein, A., Hysi, P. G., Ankra-Badu, G. A., Spector, T. D., Park, D., . . . Kayser, M. (2010). Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet, 6*(5), e1000934. doi:10.1371/journal.pgen.1000934

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S., & Hirschhorn, J. N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics, 33*(2), 177-182. doi:10.1038/ng1071

Luciano, M., Hansell, N. K., Lahti, J., Davies, G., Medland, S. E., Raikkonen, K., . . . Deary, I. J. (2011). Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biological Psychology, 86*(3), 193-202. doi:10.1016/j.biopsycho.2010.11.008

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., . . . Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res, 45*(D1), D896-D901. doi:10.1093/nar/gkw1133

Madrigal, L., & Kelly, W. (2007). Human skin-color sexual dimorphism: a test of the sexual selection hypothesis. *Am J Phys Anthropol, 132*(3), 470-482. doi:10.1002/ajpa.20453

Malnic, B., Godfrey, P. A., & Buck, L. B. (2004). The human olfactory receptor gene family (vol 101, pg 2584, 2004). *Proceedings of the National Academy of Sciences of the United States of America, 101*(18), 7205-7205. doi:10.1073/pnas.0401411101

Manley, K. J. (2015). Taste genetics and gastrointestinal symptoms experienced in chronic kidney disease. *Eur J Clin Nutr, 69*(7), 781-785. doi:10.1038/ejcn.2015.80

Manolio, T. A. (2013). Bringing genome-wide association findings into clinical use. *Nature Reviews Genetics, 14*(8), 549-558. doi:10.1038/nrg3523

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753. doi:10.1038/nature08494

Margolskee, R. F. (2002). Molecular mechanisms of bitter and sweet taste transduction. *Journal of Biological Chemistry, 277*(1), 1-4. doi:10.1074/jbc.R100054200

Maronas, O., Sochtig, J., Ruiz, Y., Phillips, C., Carracedo, A., & Lareu, M. V. (2015). The genetics of skin, hair, and eye color variation and its relevance to forensic pigmentation predictive tests. *Forensic Sci Rev, 27*(1), 13-40.

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., . . . Scherer, S. W. (2008). Structural variation of chromosomes in autism spectrum disorder. *American Journal of Human Genetics, 82*(2), 477-488. doi:10.1016/j.ajhg.2007.12.009

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., . . . Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science, 337*(6099), 1190-1195. doi:10.1126/science.1222794

McCarroll, S. A. (2008). Extending genome-wide association studies to copy-number variation. *Hum Mol Genet, 17*(R2), R135-142. doi:10.1093/hmg/ddn282

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics, 9*(5), 356-369. doi:10.1038/nrg2344

McCarthy, M. I., & Hirschhorn, J. N. (2008). Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet, 17*(R2), R156-165. doi:10.1093/hmg/ddn289

McClellan, J. M., Susser, E., & King, M. C. (2007). Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry, 190*, 194-199. doi:10.1192/bjp.bp.106.025585

McIntyre, J. C., Bose, S. C., & McClintock, T. S. (2009). Emx2 Stimulates Odorant Receptor Gene Expression and Controls OSN Axon Growth. *Chemical Senses, 34*(7), A18-A18.

McRae, J. F., Jaeger, S. R., Bava, C. M., Beresford, M. K., Hunter, D., Jia, Y. L., . . . Newcomb, R. D. (2013). Identification of Regions Associated with Variation in Sensitivity to Food-Related Odors in the Human Genome. *Current Biology, 23*(16), 1596-1600. doi:10.1016/j.cub.2013.07.031

Menini, A. (1999). Calcium signalling and regulation in olfactory neurons. *Current Opinion in Neurobiology, 9*(4), 419-426. doi:Doi 10.1016/S0959-4388(99)80063-4

Menini, A., Lagostena, L., & Boccaccio, A. (2004). Olfaction: From odorant molecules to the olfactory cortex. *News in Physiological Sciences, 19*, 101-104. doi:10.1152/nips.1507.2003

Merkulova, M., Paunescu, T. G., Azroyan, A., Marshansky, V., Breton, S., & Brown, D. (2015). Mapping the H+ (V)-ATPase interactome: identification of proteins involved in trafficking, folding, assembly and phosphorylation. *Scientific Reports, 5*. doi:ARTN 14827 10.1038/srep14827

Milinski, M., & Wedekind, C. (2001). Evidence for MHC-correlated perfume preferences in humans. *Behavioral Ecology, 12*(2), 140-149. doi:DOI 10.1093/beheco/12.2.140

Nan, H., Kraft, P., Qureshi, A. A., Guo, Q., Chen, C., Hankinson, S. E., . . . Han, J. (2009). Genome-wide association study of tanning phenotype in a population of European ancestry. *J Invest Dermatol, 129*(9), 2250-2257. doi:10.1038/jid.2009.62

Nelson, G., Chandrashekar, J., Hoon, M. A., Feng, L. X., Zhao, G., Ryba, N. J. P., & Zuker, C. S. (2002). An amino-acid taste receptor. *Nature, 416*(6877), 199-202. doi:DOI 10.1038/nature726

Nelson, G., Hoon, M. A., Chandrashekar, J., Zhang, Y. F., Ryba, N. J. P., & Zuker, C. S. (2001). Mammalian sweet taste receptors. *Cell, 106*(3), 381-390. doi:Doi 10.1016/S0092-8674(01)00451-2

Newton, J. M., Cohen-Barak, O., Hagiwara, N., Gardner, J. M., Davisson, M. T., King, R. A., & Brilliant, M. H. (2001). Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4. *Am J Hum Genet, 69*(5), 981-988. doi:10.1086/324340

Nishimura, E. K., Granter, S. R., & Fisher, D. E. (2005). Mechanisms of hair graying: incomplete melanocyte stem cell maintenance in the niche. *Science, 307*(5710), 720-724. doi:10.1126/science.1099593

Nishimura, E. K., Suzuki, M., Igras, V., Du, J., Lonning, S., Miyachi, Y., . . . Fisher, D. E. (2010). Key roles for transforming growth factor beta in melanocyte stem cell maintenance. *Cell Stem Cell, 6*(2), 130-140. doi:10.1016/j.stem.2009.12.010

Oakley, B., Chu, J. S., & Jones, L. B. (1981). Axonal-Transport Maintains Taste Responses. *Brain Research, 221*(2), 289-298. doi:Doi 10.1016/0006-8993(81)90778-2

Ober, C., Weitkamp, L. R., Cox, N., Dytch, H., Kostyu, D., & Elias, S. (1997). HLA and mate choice in humans. *Am J Hum Genet, 61*(3), 497-504. doi:10.1086/515511

Oganesian, A., Poot, M., Daum, G., Coats, S. A., Wright, M. B., Seifert, R. A., & Bowen-Pope, D. F. (2003). Protein tyrosine phosphatase RQ is a phosphatidylinositol phosphatase that can regulate cell survival and proliferation. *Proceedings of the National Academy of Sciences of the United States of America, 100*(13), 7563-7568. doi:10.1073/pnas.1336511100

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., . . . Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform, 15*(2), 256-278. doi:10.1093/bib/bbs086

Pallister, T., Sharafi, M., Lachance, G., Pirastu, N., Mohney, R. P., MacGregor, A., . . . Menni, C. (2015). Food Preference Patterns in a UK Twin Cohort. *Twin Research and Human Genetics, 18*(6), 793-805. doi:10.1017/thg.2015.69

Pan, J. T., Tian, Y., Lookingland, K. J., & Moore, K. E. (1992). Neurotensin-induced activation of hypothalamic dopaminergic neurons is accompanied by a decrease in pituitary secretion of prolactin and alpha-melanocyte-stimulating hormone. *Life Sci, 50*(25), 2011-2017.

Panhard, S., Lozano, I., & Loussouarn, G. (2012). Greying of the human hair: a worldwide survey, revisiting the '50' rule of thumb. *Br J Dermatol, 167*(4), 865-873. doi:10.1111/j.1365-2133.2012.11095.x

Panzanelli, P., Fritschy, J. M., Yanagawa, Y., Obata, K., & Sassoe-Pognetto, M. (2007). GABAergic phenotype of periglomerular cells in the rodent olfactory bulb. *J Comp Neurol, 502*(6), 990-1002. doi:10.1002/cne.21356

Parsons, B. L., & Heflich, R. H. (1997). Genotypic selection methods for the direct analysis of point mutations. *Mutat Res, 387*(2), 97-121.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *Plos Genetics, 2*(12), 2074-2093. doi:ARTN e190 10.1371/journal.pgen.0020190

Paunescu, T. G., Rodriguez, S., Benz, E., McKee, M., Tyszkowski, R., Albers, M. W., & Brown, D. (2012). Loss of the V-ATPase B1 Subunit Isoform Expressed in Non-Neuronal Cells of the Mouse Olfactory Epithelium Impairs Olfactory Function. *Plos One, 7*(9). doi:ARTN e45395 10.1371/journal.pone.0045395

Picardo, M., Dell'Anna, M. L., Ezzedine, K., Hamzavi, I., Harris, J. E., Parsad, D., & Taieb, A. (2015). Vitiligo. *Nat Rev Dis Primers, 1*, 15011. doi:10.1038/nrdp.2015.11

Pirastu, N., Kooyman, M., Traglia, M., Robino, A., Willems, S. M., Pistis, G., . . . Gasparini, P. (2016). A Genome-Wide Association Study in isolated populations reveals new genes associated to common food likings. *Rev Endocr Metab Disord, 17*(2), 209-219. doi:10.1007/s11154-016-9354-3

Pirastu, N., Kooyman, M., Traglia, M., Robino, A., Willems, S. M., Pistis, G., . . . Gasparini, P. (2015). Genome-wide association analysis on five isolated populations identifies variants of the HLA-DOA gene associated with

white wine liking. *European Journal of Human Genetics, 23*(12), 1717-1722. doi:10.1038/ejhg.2015.34

Plachez, C., Cato, K., McLeay, R. C., Heng, Y. H. E., Bailey, T. L., Gronostasjki, R. M., . . . Piper, M. (2012). Expression of Nuclear Factor One A and -B in the Olfactory Bulb. *Journal of Comparative Neurology, 520*(14), 3135-3149. doi:10.1002/cne.23081

Poliandri, A. H. B., Gamsby, J. J., Christian, M., Spinella, M. J., Loros, J. J., Dunlap, J. C., & Parker, M. G. (2011). Modulation of Clock Gene Expression by the Transcriptional Coregulator Receptor Interacting Protein 140 (RIP140). *Journal of Biological Rhythms, 26*(3), 187-199. doi:10.1177/0748730411401579

Praetorius, C., Grill, C., Stacey, S. N., Metcalf, A. M., Gorkin, D. U., Robinson, K. C., . . . Steingrimsson, E. (2013). A Polymorphism in IRF4 Affects Human Pigmentation through a Tyrosinase-Dependent MITF/TFAP2A Pathway. *Cell, 155*(5), 1022-1033. doi:10.1016/j.cell.2013.10.022

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., . . . Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics, 26*(18), 2336-2337. doi:10.1093/bioinformatics/btq419

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81*(3), 559-575. doi:10.1086/519795

Quiros, P. M., Langer, T., & Lopez-Otin, C. (2015). New roles for mitochondrial proteases in health, ageing and disease. *Nature Reviews Molecular Cell Biology, 16*(6), 345-359. doi:10.1038/nrm3984

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., . . . Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature, 444*(7118), 444-454. doi:10.1038/nature05329

Reed, D. R., Li, S., Li, X., Huang, L., Tordoff, M. G., Starling-Roney, R., . . . Bachmanov, A. A. (2004). Polymorphisms in the taste receptor gene (Tas1r3) region are associated with saccharin preference in 30 mouse strains. *Journal of Neuroscience, 24*(4), 938-946. doi:10.1523/Jneurosci.1374-03.2004

Reed, D. R., Nanthakumar, E., North, M., Bell, C., Bartoshuk, L. M., & Price, R. A. (1999). Localization of a gene for bitter-taste perception to human chromosome 5p15. *American Journal of Human Genetics, 64*(5), 1478-1480. doi:Doi 10.1086/302367

Reed, D. R., Tanaka, T., & McDaniel, A. H. (2006). Diverse tastes: Genetics of sweet and bitter perception. *Physiology & Behavior, 88*(3), 215-226. doi:10.1016/j.physbeh.2006.05.033

Reed, D. R., Zhu, G., Breslin, P. A. S., Duke, F. F., Henders, A. K., Campbell, M. J., . . . Wright, M. J. (2010). The perception of quinine taste intensity is associated with common genetic variants in a bitter receptor cluster on chromosome 12. *Human Molecular Genetics, 19*(21), 4278-4285. doi:10.1093/hmg/ddq324

Rees, J. L. (2003). Genetics of hair and skin color. *Annual Review of Genetics, 37*, 67-90. doi:10.1146/annurev.genet.37.110801.143233

Rees, J. L. (2004). The genetics of sun sensitivity in humans. *Am J Hum Genet, 75*(5), 739-751. doi:10.1086/425285

Rees, J. L., & Harding, R. M. (2012). Understanding the evolution of human pigmentation: recent contributions from population genetics. *J Invest Dermatol, 132*(3 Pt 2), 846-853. doi:10.1038/jid.2011.358

Richter, T. A., Dvoryanchikov, G. A., Chaudhari, N., & Roper, S. D. (2004). Acid-sensitive two-pore domain potassium (K2P) channels in mouse taste buds. *Journal of Neurophysiology, 92*(3), 1928-1936. doi:10.1152/jn.00273.2004

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science, 273*(5281), 1516-1517. doi:DOI 10.1126/science.273.5281.1516

Robbins, L. S., Nadeau, J. H., Johnson, K. R., Kelly, M. A., Roselliirehfuss, L., Baack, E., . . . Cone, R. D. (1993). Pigmentation Phenotypes of Variant Extension Locus Alleles Result from Point Mutations That Alter Msh Receptor Function. *Cell, 72*(6), 827-834. doi:Doi 10.1016/0092-8674(93)90572-8

Rosenzweig, S., Yan, W. T., Dasso, M., & Spielman, A. I. (1999). Possible novel mechanism for bitter taste mediated through cGMP. *Journal of Neurophysiology, 81*(4), 1661-1665.

Rozengurt, E. (2006). Taste receptors in the gastrointestinal tract. I. Bitter taste receptors and alpha-gustducin in the mammalian gut. *Am J Physiol Gastrointest Liver Physiol, 291*(2), G171-177. doi:10.1152/ajpgi.00073.2006

Rozenkrantz, L., Zachor, D., Heller, I., Plotkin, A., Weissbrod, A., Snitz, K., . . . Sobel, N. (2015). A Mechanistic Link between Olfaction and Autism Spectrum Disorder. *Current Biology, 25*(14), 1904-1910. doi:10.1016/j.cub.2015.05.048

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., . . . Grp, I. S. M. W. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*(6822), 928-933. doi:Doi 10.1038/35057149

Sano, K., Tanihara, H., Heimark, R. L., Obata, S., Davidson, M., Stjohn, T., . . . Suzuki, S. (1993). Protocadherins - a Large Family of Cadherin-Related Molecules in Central-Nervous-System. *Embo Journal, 12*(6), 2249-2256.

Sardella, M., Navone, F., Rocchi, M., Rubartelli, A., Viggiano, L., Vignali, G., . . . Cabibbo, A. (1998). KIF3C, a novel member of the kinesin superfamily: Sequence, expression, and mapping to human chromosome 2 at 2p23. *Genomics, 47*(3), 405-408. doi:DOI 10.1006/geno.1997.5123

Schmuth, M., Ortegon, A. M., Man, M. Q., Elias, P. M., Feingold, K. R., & Stahl, A. (2005). Differential expression of fatty acid transport proteins in epidermis and skin appendages. *Journal of Investigative Dermatology, 125*(6), 1174-1181. doi:10.1111/j.0022-202X.2005.23934.x

Shah, A. S., Ben-Shahar, Y., Moninger, T. O., Kline, J. N., & Welsh, M. J. (2009). Motile Cilia of Human Airway Epithelia Are Chemosensory. *Science, 325*(5944), 1131-1134. doi:10.1126/science.1173869

Sham, P. C., & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet, 15*(5), 335-346. doi:10.1038/nrg3706

Shao, Y., Cuccaro, M. L., Hauser, E. R., Raiford, K. L., Menold, M. M., Wolpert, C. M., . . . Pericak-Vance, M. A. (2003). Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. *Am J Hum Genet, 72*(3), 539-548. doi:10.1086/367846

Shapiro, T. A., Fahey, J. W., Wade, K. L., Stephenson, K. K., & Talalay, P. (2001). Chemoprotective glucosinolates and isothiocyanates of broccoli sprouts: metabolism and excretion in humans. *Cancer Epidemiol Biomarkers Prev, 10*(5), 501-508.

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*(1), 17-27. doi:10.1002/(SICI)1096-8644(200005)112:1<17::AID-AJPA3>3.0.CO;2-D

Slatkin, M. (2008). Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics, 9*(6), 477-485. doi:10.1038/nrg2361

Slominski, A., Wortsman, J., Plonka, P. M., Schallreuter, K. U., Paus, R., & Tobin, D. J. (2005). Hair follicle pigmentation. *J Invest Dermatol, 124*(1), 13-21. doi:10.1111/j.0022-202X.2004.23528.x

Smith, A. D., Fildes, A., Cooke, L., Herle, M., Shakeshaft, N., Plomin, R., & Llewellyn, C. (2016). Genetic and environmental influences on food preferences in adolescence. *Am J Clin Nutr, 104*(2), 446-453. doi:10.3945/ajcn.116.133983

Smith, J., Larue, L., & Gillespie, D. A. (2013). Chk1 is essential for the development of murine epidermal melanocytes. *Pigment Cell Melanoma Res, 26*(4), 580-585. doi:10.1111/pcmr.12100

Spencer, C. C., Su, Z., Donnelly, P., & Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet, 5*(5), e1000477. doi:10.1371/journal.pgen.1000477

Stankiewicz, P., & Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annu Rev Med, 61*, 437-455. doi:10.1146/annurev-med-100708-204735

Stark, Z., Bruno, D. L., Mountford, H., Lockhart, P. J., & Amor, D. J. (2010). De novo 325 kb microdeletion in chromosome band 10q25.3 including ATRNL1 in a boy with cognitive impairment, autism and dysmorphic features. *European Journal of Medical Genetics, 53*(5), 337-339. doi:10.1016/j.ejmg.2010.07.009

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O. P. H., Ingason, A., Steinberg, S., . . . Group. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature, 455*(7210), 232-U261. doi:10.1038/nature07229

Steingrimsson, E., Copeland, N. G., & Jenkins, N. A. (2005). Melanocyte stem cell maintenance and hair graying. *Cell, 121*(1), 9-12. doi:10.1016/j.cell.2005.03.021

Stevens, D. R., Seifert, R., Bufe, B., Muller, F., Kremmer, E., Gauss, R., . . . Lindemann, B. (2001). Hyperpolarization-activated channels HCN1 and HCN4 mediate responses to sour stimuli. *Nature, 413*(6856), 631-635. doi:Doi 10.1038/35098087

Stokowski, R. P., Pant, P. V. K., Dadd, T., Fereday, A., Hinds, D. A., Jarman, C., . . . Cox, D. R. (2007). A genomewide association study of skin pigmentation in a South Asian population. *American Journal of Human Genetics, 81*(6), 1119-1132. doi:10.1086/522235

Strand, A. D., Aragaki, A. K., Baquet, Z. C., Hodges, A., Cunningham, P., Holmans, P., . . . Olson, J. M. (2007). Conservation of regional gene expression in mouse and human brain. *Plos Genetics, 3*(4). doi:ARTN e59 10.1371/journal.pgen.0030059

Sturm, R. A. (2009). Molecular genetics of human pigmentation diversity. *Human Molecular Genetics, 18*, R9-R17. doi:10.1093/hmg/ddp003

Sturm, R. A., Teasdale, R. D., & Box, N. F. (2001). Human pigmentation genes: identification, structure and consequences of polymorphic variation. *Gene, 277*(1-2), 49-62.

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Jakobsdottir, M., . . . Stefansson, K. (2008). Two newly identified genetic determinants of pigmentation in Europeans. *Nature Genetics, 40*(7), 835-837. doi:10.1038/ng.160

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., . . . Stefansson, K. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics, 39*(12), 1443-1452. doi:10.1038/ng.2007.13

Tanner, S. M., Austin, J. L., Leone, G., Rush, L. J., Plass, C., Heinonen, K., . . . de la Chapelle, A. (2001). BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia. *Proceedings of the National Academy of Sciences of the United States of America, 98*(24), 13901-13906. doi:DOI 10.1073/pnas.241525498

Tepper, B. J. (1999). Does genetic taste sensitivity to PROP influence food preferences and body weight? *Appetite, 32*(3), 422-422. doi:DOI 10.1006/appe.1999.0240

Teucher, B., Skinner, J., Skidmore, P. M., Cassidy, A., Fairweather-Tait, S. J., Hooper, L., . . . MacGregor, A. J. (2007). Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res Hum Genet, 10*(5), 734-748. doi:10.1375/twin.10.5.734

Thapar, A., & Cooper, M. (2013). Copy number variation: what is it and what has it told us about child psychiatric disorders? *J Am Acad Child Adolesc Psychiatry, 52*(8), 772-774. doi:10.1016/j.jaac.2013.05.013

Tobin, D. J., & Paus, R. (2001). Graying: gerontobiology of the hair follicle pigmentary unit. *Exp Gerontol, 36*(1), 29-54.

Tomlinson, I. P., & Bodmer, W. F. (1995). The HLA system and the analysis of multifactorial genetic disease. *Trends Genet, 11*(12), 493-498.

Tsai, F. J., Yang, C. F., Chen, C. C., Chuang, L. M., Lu, C. H., Chang, C. T., . . . Wu, J. Y. (2010). A Genome-Wide Association Study Identifies Susceptibility Variants for Type 2 Diabetes in Han Chinese. *Plos Genetics, 6*(2). doi:ARTN e1000847 10.1371/journal.pgen.1000847

Tucker, E. S., Lehtinen, M. K., Maynard, T., Zirlinger, M., Dulac, C., Rawson, N., . . . LaMantia, A. S. (2010). Proliferative and transcriptional identity of

distinct classes of neural precursors in the mammalian olfactory epithelium. *Development, 137*(15), 2471-2481. doi:10.1242/dev.049718

Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., . . . Ritchie, M. D. (2011). Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet, Chapter 1*, Unit1 19. doi:10.1002/0471142905.hg0119s68

Turner, S. D. (2014). qqman: an R package for visualizing GWAS results usingQ-Q and manhattan plots. doi:https://doi.org/10.1101/005165

Valverde, P., Healy, E., Jackson, I., Rees, J. L., & Thody, A. J. (1995). Variants of the Melanocyte-Stimulating Hormone-Receptor Gene Are Associated with Red Hair and Fair Skin in Humans. *Nature Genetics, 11*(3), 328-330. doi:DOI 10.1038/ng1195-328

Vandenbeuch, A., Anderson, C. B., Parnes, J., Enjyoji, K., Robson, S. C., Finger, T. E., & Kinnamon, S. C. (2013). Role of the ectonucleotidase NTPDase2 in taste bud function. *Proceedings of the National Academy of Sciences of the United States of America, 110*(36), 14789-14794. doi:10.1073/pnas.1309468110

Ventura, A. K., & Worobey, J. (2013). Early influences on the development of food preferences. *Curr Biol, 23*(9), R401-408. doi:10.1016/j.cub.2013.02.037

Visscher, P. M. (2008). Sizing up human height variation. *Nat Genet, 40*(5), 489-490. doi:10.1038/ng0508-489

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five Years of GWAS Discovery. *American Journal of Human Genetics, 90*(1), 7-24. doi:10.1016/j.ajhg.2011.11.029

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet, 101*(1), 5-22. doi:10.1016/j.ajhg.2017.06.005

Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., & Kayser, M. (2011). IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci Int Genet, 5*(3), 170-180. doi:10.1016/j.fsigen.2010.02.004

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., . . . Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet, 7*(1), 98-115. doi:10.1016/j.fsigen.2012.07.005

Wang, H., Zhou, M. L., Brand, J., & Huang, L. Q. (2009). Inflammation and Taste Disorders Mechanisms in Taste Buds. *International Symposium on Olfaction and Taste, 1170*, 596-603. doi:10.1111/j.1749-6632.2009.04480.x

Wang, Q., Hasan, G., & Pikielny, C. W. (1999). Preferential expression of biotransformation enzymes in the olfactory organs of Drosophila melanogaster, the antennae. *Journal of Biological Chemistry, 274*(15), 10309-10315. doi:DOI 10.1074/jbc.274.15.10309

Weale, M. E. (2010). Quality Control for Genome-Wide Association Studies. *Genetic Variation: Methods and Protocols, 628*, 341-372. doi:10.1007/978-1-60327-367-1_19

Wehrle-Haller, B. (2003). The role of Kit-ligand in melanocyte development and epidermal homeostasis. *Pigment Cell Res, 16*(3), 287-296.

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics, 76*(5), 887-893. doi:Doi 10.1086/429864

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterlander, M., . . . Burger, J. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc Natl Acad Sci U S A, 111*(13), 4832-4837. doi:10.1073/pnas.1316513111

Winney, B., Boumertit, A., Day, T., Davison, D., Echeta, C., Evseeva, I., . . . Bodmer, W. (2012). People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population. *European Journal of Human Genetics, 20*(2), 203-210. doi:10.1038/ejhg.2011.127

Wise, A. L., Gyi, L., & Manolio, T. A. (2013). eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet, 92*(5), 643-647. doi:10.1016/j.ajhg.2013.03.017

Wittke-Thompson, J. K., Pluzhnikov, A., & Cox, N. J. (2005). Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet, 76*(6), 967-986. doi:10.1086/430507

Wolf Horrell, E. M., Boulanger, M. C., & D'Orazio, J. A. (2016). Melanocortin 1 Receptor: Structure, Function, and Regulation. *Front Genet, 7*, 95. doi:10.3389/fgene.2016.00095

Wooding, S. (2006). Phenylthiocarbamide: A 75-year adventure in genetics and natural selection. *Genetics, 172*(4), 2015-2023.

Wooding, S., Kim, U. K., Bamshad, M. J., Larsen, J., Jorde, L. B., & Drayna, D. (2004). Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *Am J Hum Genet, 74*(4), 637-646. doi:10.1086/383092

Yang, Z. H., & Goldstein, L. S. B. (1998). Characterization of the KIF3C neural kinesin-like motor from mouse. *Molecular Biology of the Cell, 9*(2), 249-261.

Yeh, T. C., Ogawa, W., Danielsen, A. G., & Roth, R. A. (1996). Characterization and cloning of a 58/53-kDa substrate of the insulin receptor tyrosine kinase. *Journal of Biological Chemistry, 271*(6), 2921-2928.

Yoshitane, H., Ozaki, H., Terajima, H., Du, N. H., Suzuki, Y., Fujimori, T., . . . Fukada, Y. (2014). CLOCK-Controlled Polyphonic Regulation of Circadian Rhythms through Canonical and Noncanonical E-Boxes. *Molecular and Cellular Biology, 34*(10), 1776-1787. doi:10.1128/Mcb.01465-13

Yu, H., Wang, L. F., Lv, L. X., Ma, C. C., Du, B., Lu, T. L., . . . Yue, W. H. (2016). Genome-Wide Association Study Suggested the PTPRD Polymorphisms Were Associated With Weight Gain Effects of Atypical Antipsychotic Medications. *Schizophrenia Bulletin, 42*(3), 814-823. doi:10.1093/schbul/sbv179

Yu, Y., Ma, J., Xiao, Y. B., Yang, Q. J., Kang, H. L., Zhen, J., . . . Chen, L. (2015). KLF15 Is an Essential Negative Regulatory Factor for the Cardiac Remodeling Response to Pressure Overload. *Cardiology, 130*(3), 143-152. doi:10.1159/000369382

Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nat Rev Genet, 16*(3), 172-183. doi:10.1038/nrg3871

Zhang, K. X., Huentelman, M. J., Rao, F. W., Sun, E. I., Corneveaux, J. J., Schork, A. J., . . . Geno, I. C. B. P. (2014). Genetic Implication of a Novel Thiamine

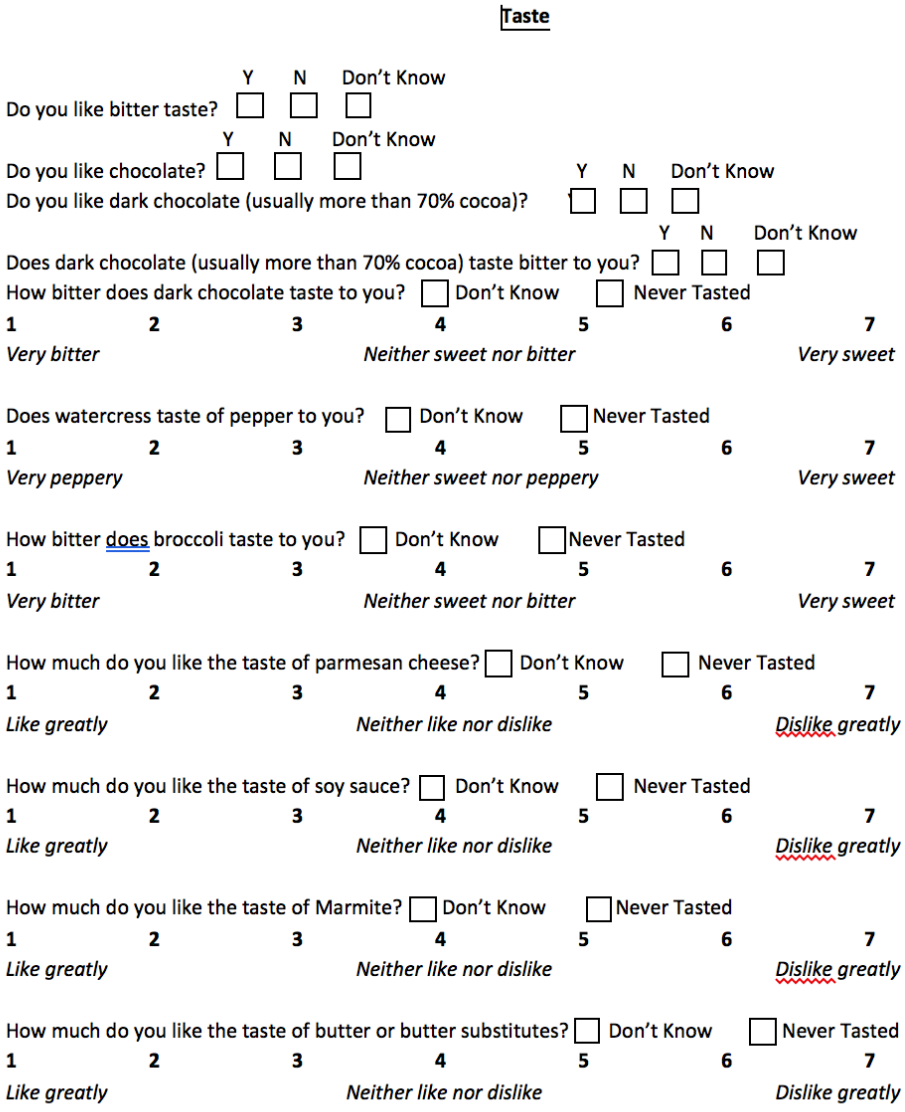
Transporter in Human Hypertension. *Journal of the American College of Cardiology, 63*(15), 1542-1555. doi:10.1016/j.jacc.2014.01.007

Zhang, M., Song, F., Liang, L., Nan, H., Zhang, J., Liu, H., . . . Han, J. (2013). Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum Mol Genet, 22*(14), 2948-2959. doi:10.1093/hmg/ddt142

Zhao, G. Q., Zhang, Y., Hoon, M. A., Chandrashekar, J., Erlenbach, I., Ryba, N. J., & Zuker, C. S. (2003). The receptors for mammalian sweet and umami taste. *Cell, 115*(3), 255-266.

9 Appendix

A.1 | A part of the questionnaire form used in the PoBI project for taste and food preference data collection.



Taste

Do you like bitter taste? Y ☐ N ☐ Don't Know ☐

Do you like chocolate? Y ☐ N ☐ Don't Know ☐

Do you like dark chocolate (usually more than 70% cocoa)? Y ☐ N ☐ Don't Know ☐

Does dark chocolate (usually more than 70% cocoa) taste bitter to you? Y ☐ N ☐ Don't Know ☐

How bitter does dark chocolate taste to you? ☐ Don't Know ☐ Never Tasted

1 **2** **3** **4** **5** **6** **7**

Very bitter *Neither sweet nor bitter* *Very sweet*

Does watercress taste of pepper to you? ☐ Don't Know ☐ Never Tasted

1 **2** **3** **4** **5** **6** **7**

Very peppery *Neither sweet nor peppery* *Very sweet*

How bitter does broccoli taste to you? ☐ Don't Know ☐ Never Tasted

1 **2** **3** **4** **5** **6** **7**

Very bitter *Neither sweet nor bitter* *Very sweet*

How much do you like the taste of parmesan cheese? ☐ Don't Know ☐ Never Tasted

1 **2** **3** **4** **5** **6** **7**

Like greatly *Neither like nor dislike* *Dislike greatly*

How much do you like the taste of soy sauce? ☐ Don't Know ☐ Never Tasted

1 **2** **3** **4** **5** **6** **7**

Like greatly *Neither like nor dislike* *Dislike greatly*

How much do you like the taste of Marmite? ☐ Don't Know ☐ Never Tasted

1 **2** **3** **4** **5** **6** **7**

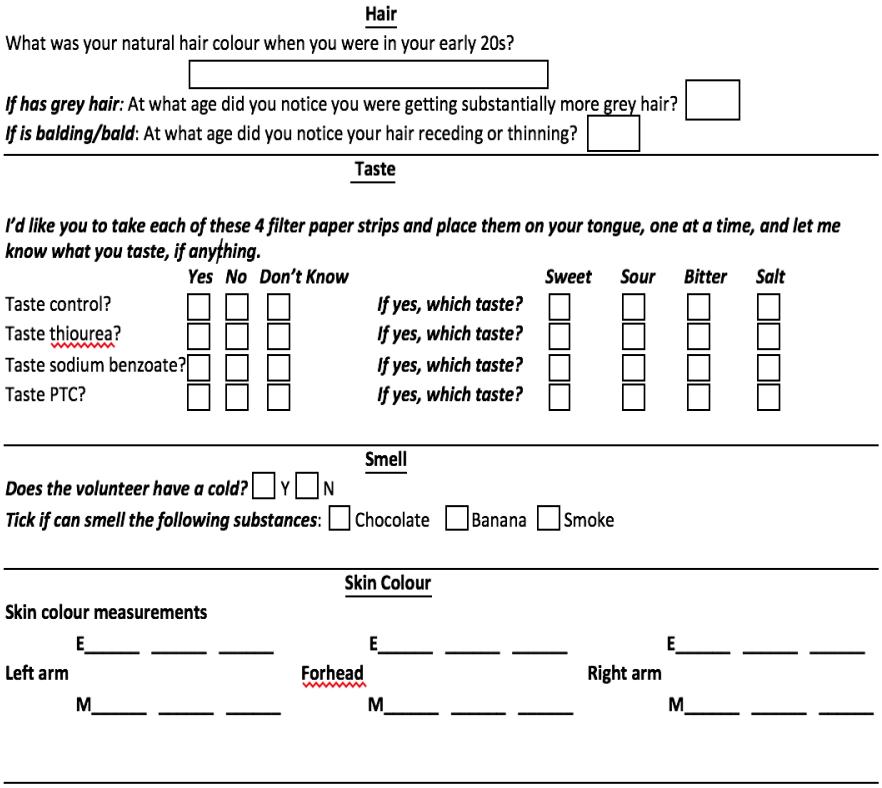
Like greatly *Neither like nor dislike* *Dislike greatly*

How much do you like the taste of butter or butter substitutes? ☐ Don't Know ☐ Never Tasted

1 **2** **3** **4** **5** **6** **7**

Like greatly *Neither like nor dislike* *Dislike greatly*

A.2 | A part of the questionnaire form used in the PoBI project for hair colour, olfactory perception and skin colour data collection.



Hair

What was your natural hair colour when you were in your early 20s?

If has grey hair: At what age did you notice you were getting substantially more grey hair? ☐

If is balding/bald: At what age did you notice your hair receding or thinning? ☐

Taste

I'd like you to take each of these 4 filter paper strips and place them on your tongue, one at a time, and let me know what you taste, if anything.

	Yes	*No*	*Don't Know*		*Sweet*	*Sour*	*Bitter*	*Salt*
Taste control?	☐	☐	☐	*If yes, which taste?*	☐	☐	☐	☐
Taste thiourea?	☐	☐	☐	*If yes, which taste?*	☐	☐	☐	☐
Taste sodium benzoate?	☐	☐	☐	*If yes, which taste?*	☐	☐	☐	☐
Taste PTC?	☐	☐	☐	*If yes, which taste?*	☐	☐	☐	☐

Smell

Does the volunteer have a cold? ☐ Y ☐ N

Tick if can smell the following substances: ☐ Chocolate ☐ Banana ☐ Smoke

Skin Colour

Skin colour measurements

E_____ _____ _____	E_____ _____ _____	E_____ _____ _____
Left arm	**Forhead**	**Right arm**
M_____ _____ _____	M_____ _____ _____	M_____ _____ _____

Appendix B

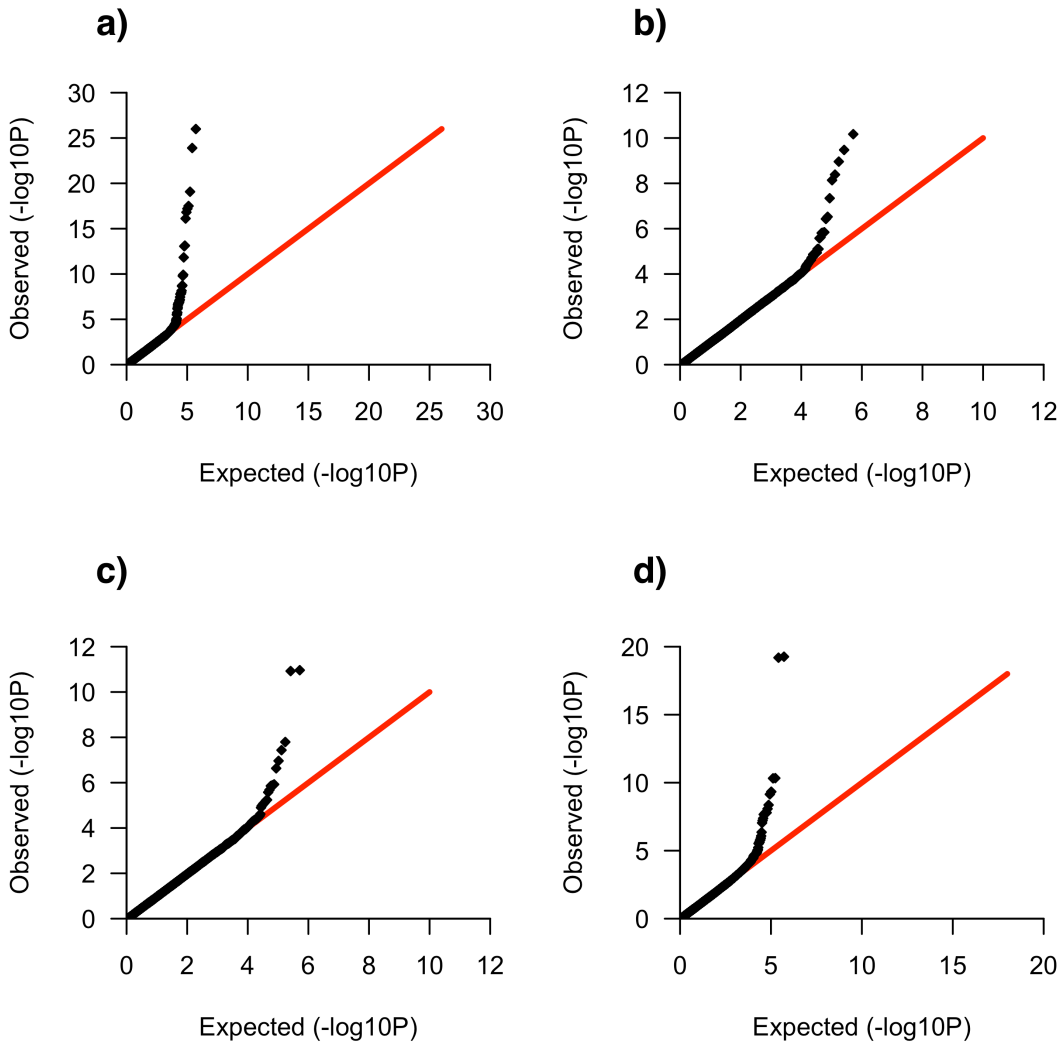
B.1 | A list of the SNPs that reached genome-wide significance with red and light hair colour traits.

Phenotype	***SNP***	***Chr***	***Gene[a]***	***P-value[b] (OR)***
Red hair	rs12918773	16	*SPATA33 / MC1R*	1.01×10^{-26} (8.17)
	rs258322	16	*CDK10 / MC1R*	1.26×10^{-24} (7.72)
	rs7195043	16	*DEF8*	8.39×10^{-20} (8.04)
	rs4785751	16	*DEF8*	3.16×10^{-18} (7.16)
	rs11648898	16	*AFG3L1P / MC1R*	6.27×10^{-18} (5.92)
	rs8049897	16	*DEF8*	1.54×10^{-17} (5.52)
	rs9939542	16	*AFG3L1P / MC1R*	7.61×10^{-17} (7.20)
	rs12924101	16	*FANCA*	7.38×10^{-14} (4.67)
	rs4785763	16	*AFG3L1P / MC1R*	8.73×10^{-14} (5.36)
	rs164741	16	*DPEP1*	1.48×10^{-12} (5.66)
	rs4238833	16	*AFG3L1P / MC1R*	1.21×10^{-10} (4.23)
	rs11076650	16	*AFG3L1P / MC1R*	1.63×10^{-10} (3.97)
	rs10852628	16	*DBNDD1*	1.81×10^{-9} (4.07)
	rs8060502	16	*SPG7*	2.11×10^{-9} (4.45)
	rs4785752	16	*AFG3L1P / MC1R*	6.99×10^{-9} (3.39)
	rs4785759	16	*AFG3L1P / MC1R*	1.06×10^{-8} (3.53)
	rs2353033	16	*ANKRD11*	1.50×10^{-8} (3.55)
	rs12932628	16	*RPL13*	1.65×10^{-8} (3.66)
	rs7195066	16	*FANCA*	1.65×10^{-8} (4.04)
	rs7188458	16	*SPATA33 / MC1R*	3.49×10^{-8} (5.07)

Light versus dark hair	rs12918773	16	*SPATA33 / MC1R*	5.45×10^{-20} (2.75)
	rs258322	16	*CDK10 / MC1R*	6.44×10^{-20} (2.64)
	rs12913832	15	*HERC2 / OCA2*	4.63×10^{-11} (2.52)
	rs7195043	16	*DEF8*	4.79×10^{-11} (2.50)
	rs4785751	16	*DEF8*	4.62×10^{-10} (2.29)
	rs12203592	6	*IRF4*	7.18×10^{-10} (2.36)
	rs3935591	15	*HERC2 / OCA2*	4.36×10^{-9} (2.74)
	rs12924101	16	*FANCA*	8.34×10^{-9} (1.98)
	rs164741	16	*DPEP1*	1.60×10^{-8} (1.91)
	rs11648898	16	*AFG3L1P / MC1R*	1.64×10^{-8} (1.89)
	rs9939542	16	*AFG3L1P / MC1R*	1.69×10^{-8} (2.78)
	rs916977	15	*HERC2 / OCA2*	2.06×10^{-8} (2.49)
	rs8049897	16	*DEF8*	2.18×10^{-8} (1.86)
	rs1667394	15	*HERC2 / OCA2*	4.20×10^{-8} (2.40)

SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; [b]Fisher's exact test; OR, Odds ratio.

B.2 | Multi-panel Q-Q plots for the GWAS of hair colour. The $-log_{10}$ of observed p-values (Fisher's exact test) for each SNP were plotted against the $-log_{10}$ of expected p-values. The red line indicates the null hypothesis of no true association. a) red versus non-red b) black versus non-black c) blond versus non-blond d) light versus dark hair.



B.3 | Distribution of rs258322 genotypes in participants with red and non-red hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Red	15	46	30	91
Non-red	12	299	1183	1494
Total	27	345	1213	1585

$P_{Dom} = 7.9 \times 10^{-20}$, $OR_{Dom} = 7.72$; $P_{Rec} = 7.6 \times 10^{-13}$, $OR_{Rec} = 24.21$

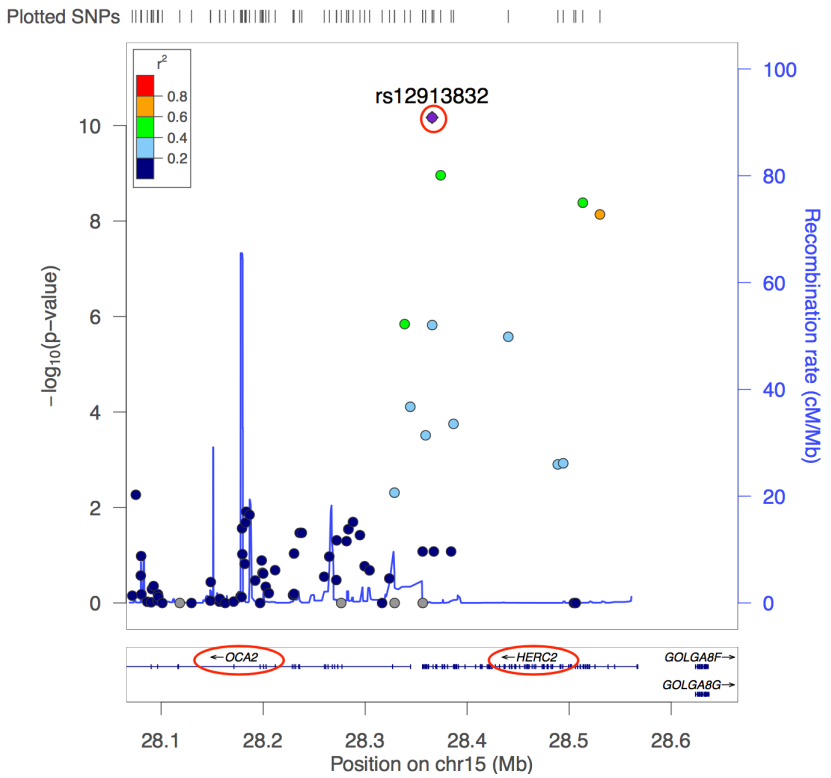
B.4 | Distribution of rs9850626 genotypes in participants with red and non-red hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Red	19	37	35	91
Non-red	76	559	857	1492
Total	95	596	892	1583

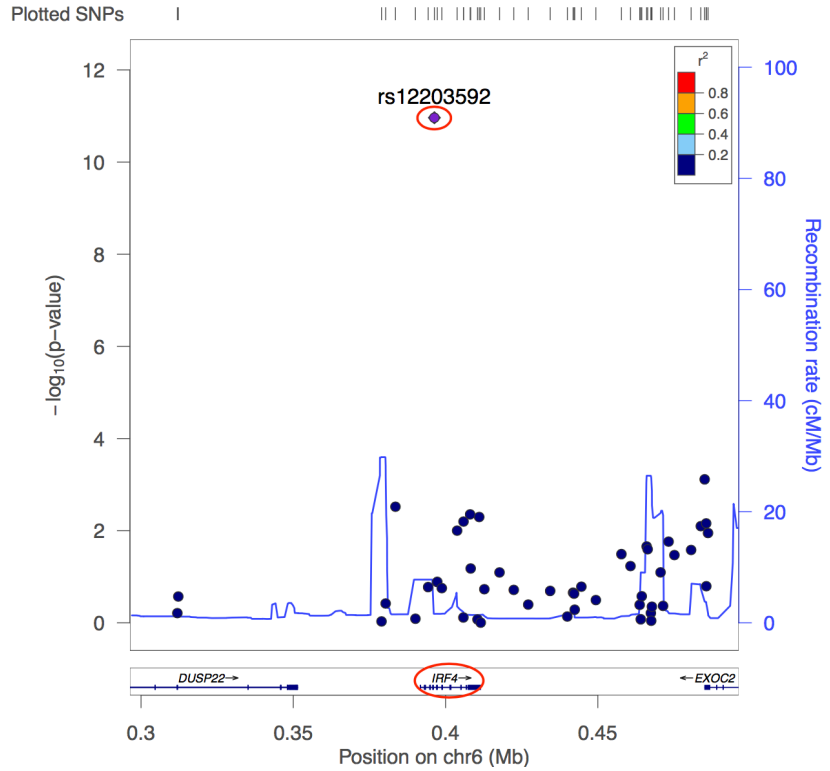
$P_{Dom} = 0.0004593$, $OR_{Dom} = 2.15$; $P_{Rec} = 5.6 \times 10^{-7}$, $OR_{Rec} = 4.90$

B.5 | Regional association plots for the GWAS of hair colour a) Black and b) Blond. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 15 and 6. The purple diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)



b)



B.6 | Distribution of rs12203592 genotypes in participants with black and non-black hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Black	11	64	38	113
Non-black	71	443	957	1471
Total	82	507	995	1584

$P_{Dom} = 1.1\times10^{-10}$, $OR_{Dom} = 3.67$; $P_{Rec} = 0.04251$, $OR_{Rec} = 2.12$

B.7 | Distribution of rs35391 genotypes in participants with black and non-black hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Black	0	18	95	113
Non-black	0	42	1429	1471
Total	0	60	1524	1584

$P_{Dom} = 4.5\times10^{-8}$, $OR_{Dom} = 6.43$, $P_{Rec} > 0.05$

B.8 | Distribution of rs12913832 genotypes in participants with blond and non-blond hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Blond	4	39	210	253
Non-blond	52	383	663	1098
Total	56	422	873	1351

$P_{Dom} = 1.5\times10^{-12}$, $OR_{Dom} = 3.20$; $P_{Rec} = 0.0219$, $OR_{Rec} = 3.09$

B.9 | Distribution of rs12913832 genotypes in participants with light and dark hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Light-hair	7	70	267	344
Dark-hair	67	443	701	1211
Total	74	513	968	1555

$P_{Dom} = 1.0\times10^{-11}$, $OR_{Dom} = 2.52$; $P_{Rec} = 0.005925$, $OR_{Rec} = 2.81$

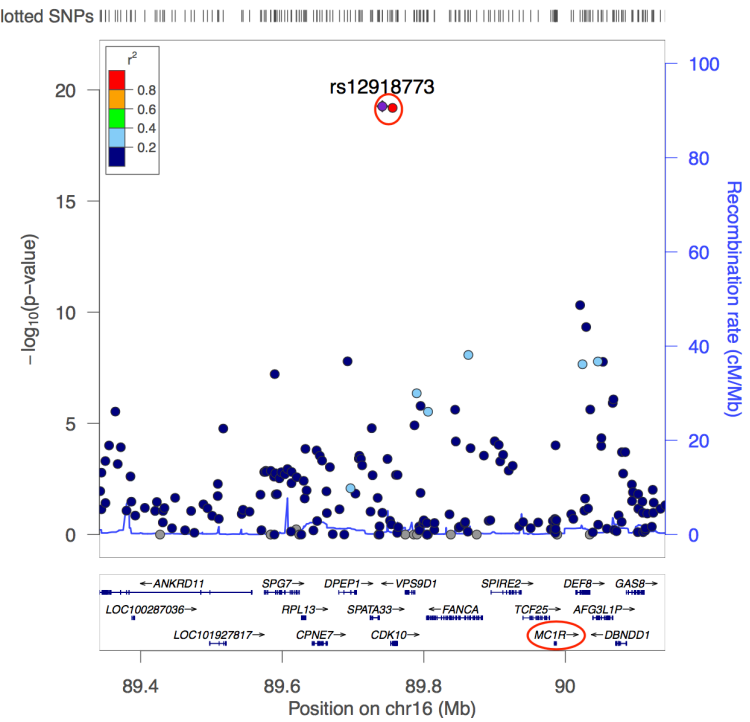
B.10 | Distribution of rs12203592 genotypes among participants with light and dark hair.

Genotype	***AA***	***Aa***	***aa***	***Total***
Light-hair	8	72	264	344
Dark-hair	74	432	705	1211
Total	82	504	969	1555

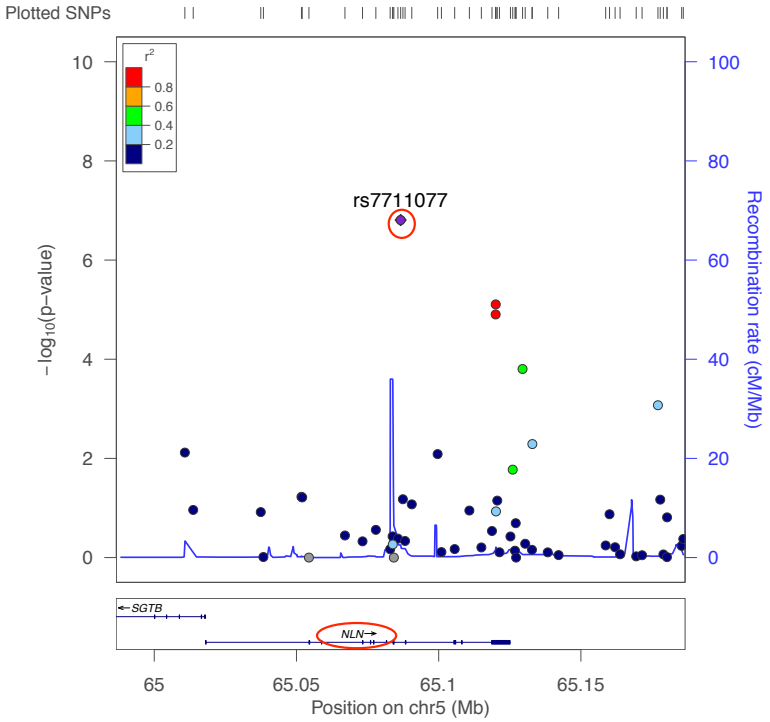
$P_{Dom} = 1.5\times10^{-10}$, $OR_{Dom} = 2.36$; $P_{Rec} = 0.003913$, $OR_{Rec} = 2.73$

B.11 | Regional association plots for the GWAS of hair colour a) light versus dark and b) grey hair. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 16 and 5. The purple diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)

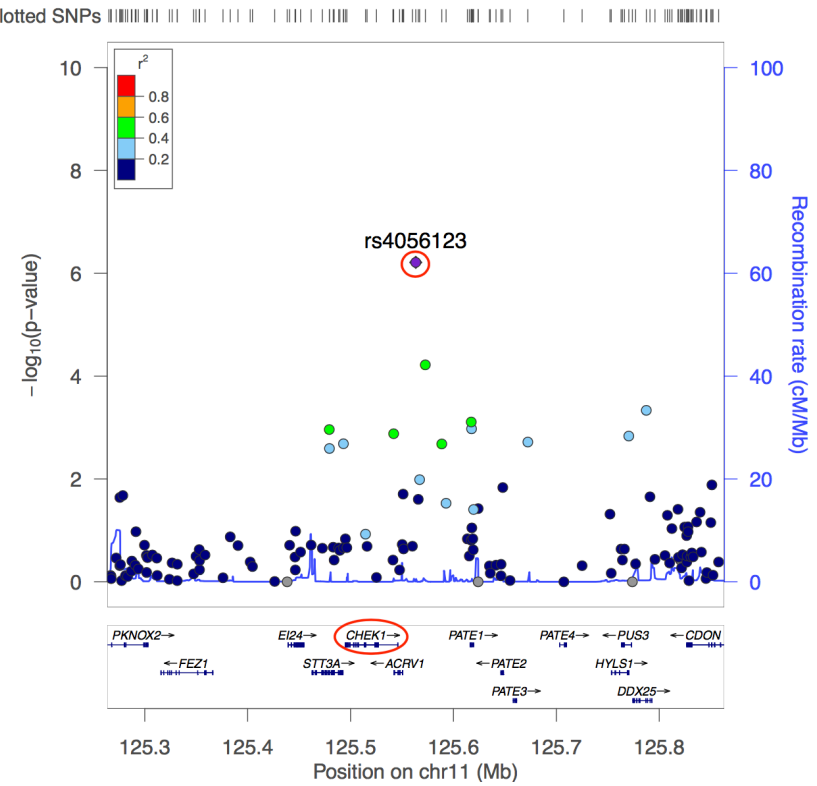


b)

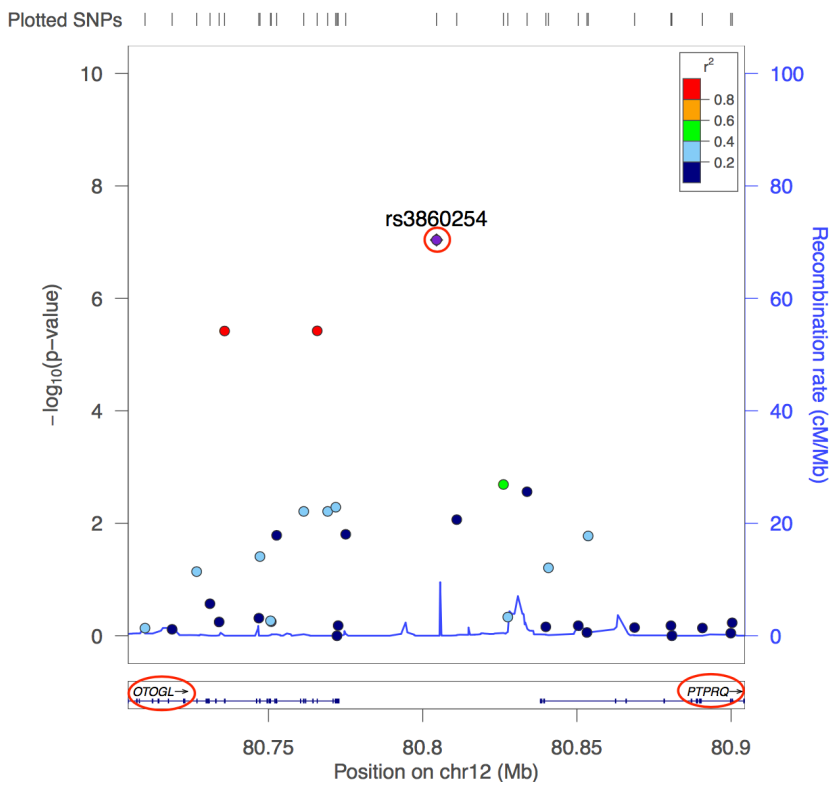


B.12 | Regional association plots for the GWAS of skin colour a) Melanin and b) Erythema. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 11 and 12. The purple diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)



b)



B.13 | Distribution of rs4056123 genotypes in participants of Melanin-index GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Melanin (low)	167	301	118	586
Melanin (High)	195	473	303	971
Total	362	774	421	1557

$P_{Dom} = 1.7\times10^{-6}$, $OR_{Dom} = 1.79$; $P_{Rec} = 0.0001582$, $OR_{Rec} = 1.58$

B.14 | Distribution of rs28777 genotypes in participants of Melanin-index GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Melanin (low)	0	14	568	582
Melanin (High)	1	61	902	964
Total	1	75	1470	1546

$P_{Dom} = 0.0002465$, $OR_{Dom} = 2.78$, $P_{Rec} > 0.05$

B.15 | Distribution of rs3860254 genotypes in participants of Erythema-index GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Erythema (low)	47	60	84	191
Erythema (High)	172	699	498	1369
Total	219	759	582	1560

$P_{Dom} = 0.04587$, $OR_{Dom} = 1.37$; $P_{Rec} = 3.0\times10^{-5}$, $OR_{Rec} = 2.26$

B.16 | Distribution of rs4341393 genotypes in participants of Erythema-index GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Erythema (low)	37	74	80	191
Erythema (High)	330	728	312	1370
Total	367	802	392	1561

$P_{Dom} = 6.0 \times 10^{-8}$, $OR_{Dom} = 2.44$; $P_{Rec} > 0.05$

Appendix C

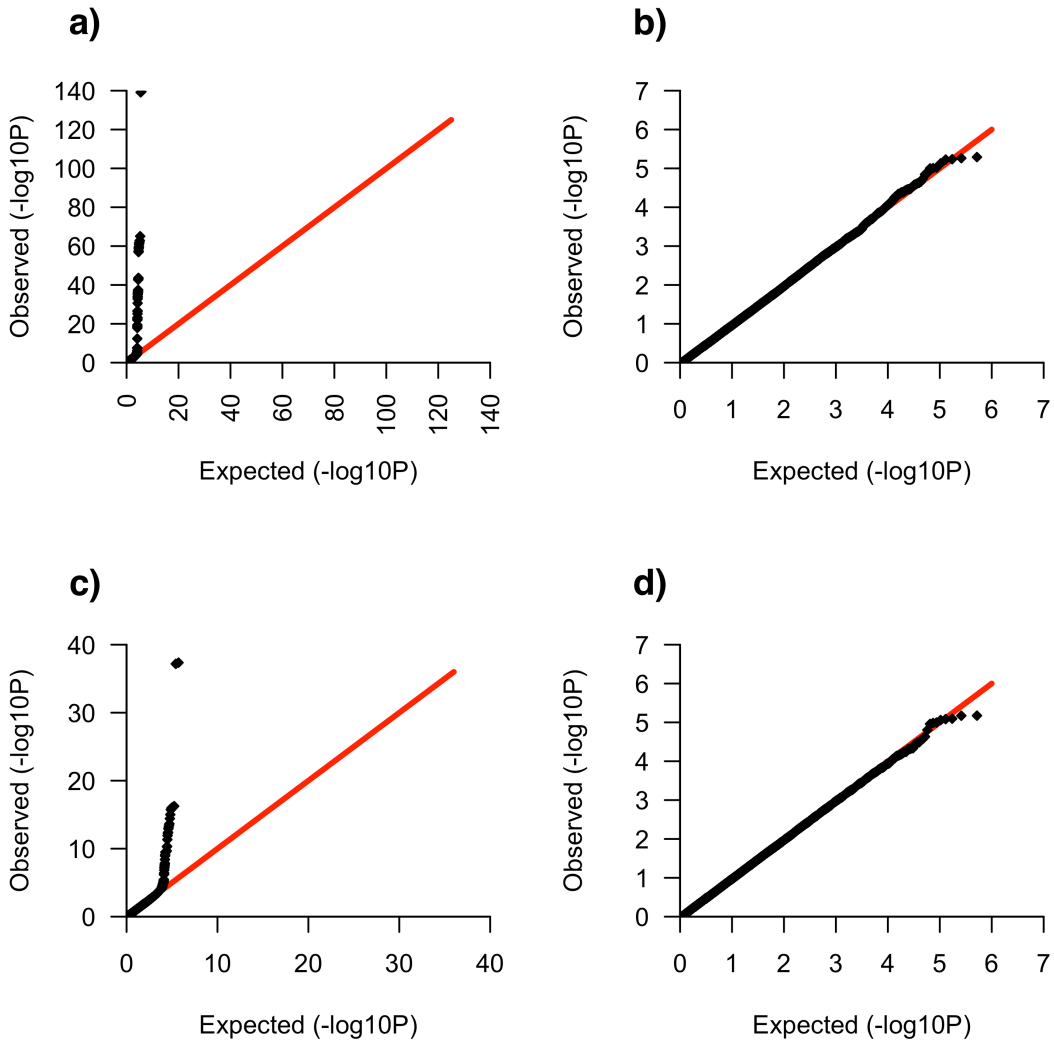
C.1 | A List of the SNPs that reached genome-wide significance for PTC and PROP taste perception.

Phenotype	***SNP***	***region***	***Gene***	***P-value***	***OR***
PTC	rs1726866	7q34	*TAS2R38*	2.75×10^{-140}	29.81
	rs10246939	7q34	*TAS2R38*	6.76×10^{-140}	29.73
	rs7794708	7q34	*MGAM*	8.48×10^{-66}	7.64
	rs10464444	7q34	*PRSS37*	1.64×10^{-63}	7.38
	rs6976028	7q34	*MGAM/OR9A4*	1.73×10^{-63}	7.33
	rs11765575	7q34	*PRSS37*	1.43×10^{-62}	7.20
	rs9640357	7q34	*PRSS37*	2.35×10^{-62}	7.18
	rs1285950	7q34	*MGAM/CLEC5A*	1.12×10^{-61}	7.06
	rs6959360	7q34	-	1.77×10^{-60}	6.94
	rs2570407	7q34	*MGAM*	4.41×10^{-60}	7.10
	rs12539499	7q34	*MGAM/CLEC5A*	2.05×10^{-59}	6.97
	rs11767947	7q34	*MGAM/OR9A4*	3.23×10^{-58}	6.84
	rs11770855	7q34	*MGAM/CLEC5A*	1.16×10^{-57}	6.76
	rs1433594	7q34	*OR9A3P*	2.28×10^{-44}	5.16
	rs1285933	7q34	*MGAM/CLEC5A*	5.87×10^{-44}	5.14
	rs745162	7q34	*MGAM/OR9A4*	1.58×10^{-43}	5.07
	rs12669721	7q34	*PRSS37*	1.85×10^{-43}	5.03
	rs13231650	7q34	-	3.40×10^{-38}	4.54
	rs6963959	7q34	*WEE2*	7.55×10^{-38}	4.48
	rs2227264	7q34	*TAS2R5/SSBP1*	3.24×10^{-37}	4.49
	rs2190245	7q34	*TAS2R5/SSBP1*	7.28×10^{-37}	4.44
	rs2234012	7q34	*TAS2R5/SSBP1*	7.83×10^{-37}	4.44
	rs12539227	7q34	*SSBP1, TAS2R4, TAS2R6P*	1.22×10^{-36}	4.43
	rs1859646	7q34	*TAS2R5*	4.22×10^{-36}	4.34

	rs2234002	7q34	*TAS2R4, SSBP1*	7.08×10^{-36}	4.31
	rs12540335	7q34	*SSBP1*	8.50×10^{-35}	4.20
	rs2270009	7q34	*TAS2R3/SSBP1*	1.69×10^{-34}	4.17
	rs765007	7q34	*TAS2R3/SSBP1*	1.53×10^{-33}	4.04
	rs6947065	7q34	*WEE2*	2.44×10^{-31}	4.02
	rs7791469	7q34	-	2.05×10^{-27}	3.69
	rs10261515	7q34	-	4.80×10^{-27}	3.72
	rs768055	7q34	*WEE2*	7.76×10^{-26}	3.98
	rs11773340	7q34	*AGK*	5.79×10^{-24}	3.25
	rs4725559	7q34	*AGK*	3.48×10^{-23}	3.20
	rs3757470	7q34	*AGK*	4.04×10^{-23}	3.20
	rs3800989	7q34	-	9.90×10^{-23}	3.12
	rs12703398	7q34	*AGK*	4.24×10^{-20}	2.91
	rs1569085	7q34	-	2.50×10^{-19}	2.83
	rs1285961	7q34	-	1.29×10^{-18}	2.61
	rs4632970	7q34	*AGK*	4.36×10^{-13}	2.93
	rs11981914	7q34	*TAS2R38*	2.27×10^{-8}	5.35
	rs2366254	7q34	-	4.23×10^{-8}	1.95
Phenotype	**SNP**	**region**	***Gene***	**P value**	**OR**
PROP	rs1726866	7q34	*TAS2R38*	4.38×10^{-38}	4.60
	rs10246939	7q34	*TAS2R38*	6.58×10^{-38}	4.59
	rs10464444	7q34	*PRSS37*	5.58×10^{-17}	2.49
	rs11765575	7q34	*PRSS37*	7.34×10^{-17}	2.47
	rs9640357	7q34	*PRSS37*	8.80×10^{-17}	2.47
	rs6976028	7q34	*MGAM/OR9A4*	1.05×10^{-16}	2.49
	rs7794708	7q34	*MGAM*	1.90×10^{-16}	2.47
	rs6959360	7q34	-	9.70×10^{-16}	2.40
	rs1285950	7q34	*MGAM/CLEC5A*	3.83×10^{-15}	1.91
	rs12539499	7q34	*MGAM/CLEC5A*	2.17×10^{-14}	2.33
	rs11767947	7q34	*MGAM/OR9A4*	3.47×10^{-14}	2.31
	rs2570407	7q34	*MGAM*	6.43×10^{-14}	2.29

	rs11770855	7q34	*MGAM/CLEC5A*	1.20×10^{-13}	2.27
	rs1285933	7q34	*MGAM/CLEC5A*	4.03×10^{-13}	2.32
	rs1433594	7q34	*OR9A3P*	5.13×10^{-13}	2.28
	rs745162	7q34	*MGAM/OR9A4*	1.17×10^{-12}	2.28
	rs12669721	7q34	*PRSS37*	4.56×10^{-12}	2.19
	rs6963959	7q34	*WEE2*	4.58×10^{-11}	2.03
	rs13231650	7q34	-	4.97×10^{-11}	2.05
	rs2227264	7q34	*TAS2R5*	2.12×10^{-10}	2.01
	rs1859646	7q34	*TAS2R5*	2.23×10^{-10}	1.98
	rs2234012	7q34	*TAS2R5*	2.63×10^{-10}	1.99
	rs12540335	7q34	*SSBP1*	2.65×10^{-10}	2.00
	rs2190245	7q34	*TAS2R4, SSBP1*	2.89×10^{-10}	1.99
	rs2234002	7q34	*TAS2R4, SSBP1*	2.97×10^{-10}	1.96
	rs2270009	7q34	*TAS2R3, SSBP1*	3.27×10^{-10}	1.98
	rs12539227	7q34	*TAS2R4*	3.36×10^{-10}	1.99
	rs765007	7q34	*TAS2R3, SSBP1*	3.38×10^{-10}	1.96
	rs768055	7q34	*WEE2*	8.24×10^{-10}	1.82
	rs6947065	7q34	*WEE2*	1.15×10^{-9}	1.96
	rs7791469	7q34	-	3.98×10^{-9}	1.85
	rs3800989	7q34	-	1.38×10^{-8}	1.85
	rs12703398	7q34	*AGK*	2.33×10^{-8}	1.84
	rs10261515	7q34	-	2.81×10^{-8}	1.92
Sodium Benzoate	rs3789421	1p22.1	*ABCA4*	6.7×10^{-6}	1.44
	rs2645075	1p36.33	*MORN1*	6.73×10^{-6}	4.16
	rs10158007	1p36.21	*KAZN*	8.07×10^{-6}	2.27
	rs10447360	6q13	*B3GAT2*	8.24×10^{-6}	1.65
	rs12141954	1p36.31	-	8.70×10^{-6}	1.94

C.2 | Q-Q plots for the GWAS of taste perception: a) PTC taste perception, b) after excluding the most significant associations with PTC taste perception, c) PROP, d) sodium benzoate. The $-log_{10}$ of observed p-values (black dots) for each SNP were plotted against the $-log_{10}$ of expected p-values. The red line indicates the null hypothesis of no true association.



C.3 | Distribution of rs4969384 genotypes in PTC tasters and non-tasters.

Genotype	***AA***	***Aa***	***aa***	***Total***
Non-tasters	25	144	352	521
Tasters	58	347	480	885
Total	83	491	832	1406

$P_{Dom} = 9.7\times10^{-7}$, $OR_{Dom} = 1.75$; $P_{Rec} > 0.05$

C.4 | Distribution of rs10210235 genotypes in PROP tasters and non-tasters.

Genotype	*AA*	*Aa*	*aa*	*Total*
Non-tasters	144	409	166	719
Tasters	156	292	228	676
Total	300	701	394	1395

$P_{Dom} = 1.0\times10^{-5}$, $OR_{Dom} = 1.69$; $P_{Rec} > 0.05$

C.5 | Distribution of rs11616269 genotypes in PROP tasters and non-tasters.

Genotype	***AA***	***Aa***	***aa***	***Total***
Non-tasters	5	191	523	719
Tasters	24	128	525	677
Total	29	319	1048	1396

$P_{Dom} = 0.04109$, $OR_{Dom} = 1.29$; $P_{Rec} = 0.0002149$, $OR_{Rec} = 5.24$

C.6 | Distribution of rs2645075 genotypes in sodium benzoate tasters and non-tasters.

Genotype	***AA***	***Aa***	***aa***	***Total***
Non-tasters	10	189	312	511
Tasters	77	337	588	1002
Total	87	526	900	1513

$P_{Dom} > 0.05$; $P_{Rec} = 1.8\times10^{-6}$, $OR_{Rec} = 4.16$

C.7 | Distribution of rs12141954 genotypes in sodium benzoate tasters and non-tasters.

Genotype	***AA***	***Aa***	***aa***	***Total***
Non-tasters	124	232	154	510
Tasters	142	507	353	1002
Total	266	739	507	1512

$P_{Dom} = 0.05041$, $OR_{Dom} = 1.25$; $P_{Rec} = 1.5\times10^{-6}$, $OR_{Rec} = 1.94$

Appendix D

D.1 | Associations of genetic variants with food preference in the PoBI data.

Phenotype	***Sample Size (like/dislike)***	***SNP***	***Chr***	***Position***	***Gene[a]***	***P-value[b] (OR)***
Butter preference	1434 (1358/76)	rs11126303 rs10976860	2 9	26173503 8227101	*KIF3C* -	2.3×10^{-6} (5.65) 8.3×10^{-6} (3.12)
Broccoli preference*	472 (176/296)	rs4442431	1	234020481	*KCNK1/SLC35F3*	2.0×10^{-6} (2.69)
		rs13241883	7	43512070	*HECW1/COA1*	4.9×10^{-6} (2.08)
		rs1519113	2	66652032	*MEIS1*	7.6×10^{-6} (5.82)
		rs1200115	1	169061558	-	7.9×10^{-6} (2.39)

		rs6806547	3	63893477	*ATXN7*	8.4×10^{-6} (2.38)
Soy sauce	1130(857/273)	rs12101368	15	100412680	-	9.7×10^{-7} (1.81)
		rs4822135	22	42876604	*RRP7A*	1.4×10^{-6} (2.23)
		rs1531846	8	90074006	-	2.4×10^{-6} (2.12)
		rs11779424	8	90158554	-	5.2×10^{-6} (2.07)
		rs386057	5	685748	*TPPP*	6.6×10^{-6} (1.82)
Parmesan cheese	1198(928/270)	rs2435200	17	44071851	*MAPT/STH*	3.0×10^{-6} (2.93)
		rs7200773	16	13845690	-	3.9×10^{-6} (2.30)
		rs8063755	16	13836569	-	4.3×10^{-6} (2.94)

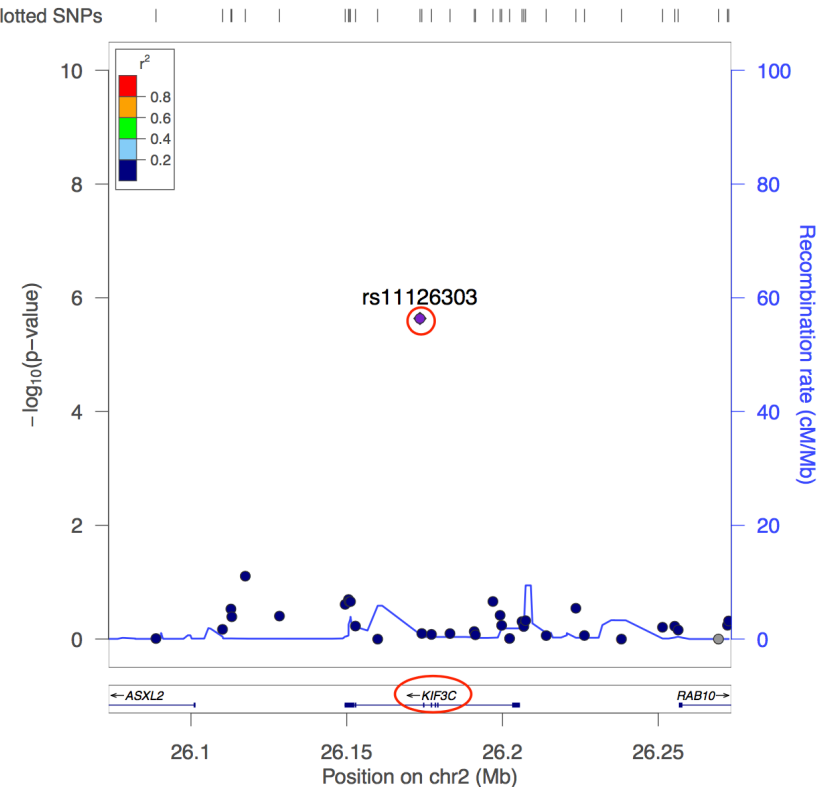
		rs10493631	1	80477880	*HMGB1P18*	6.5×10^{-06} (1.84)
		rs2184898	10	119418104	-	8.0×10^{-6} (2.93)
Marmite preference	1316 (739/577)	rs11979436	7	78108703	*MAGI2*	1.2×10^{-6} (10.89)
		rs1045643	16	4931513	*PPL/UBN1*	2.5×10^{-6} (1.93)
		rs2306660	19	45802863	*MARK4*	3.6×10^{-6} (3.17)
		rs2837480	21	41546955	*DSCAM*	9.0×10^{-6} (2.09)
		rs6974799	7	78117090	*MAGI2*	9.4×10^{-6} (1.68)
Watercress*	1068 (999/69)	rs1560160	5	66473137	*CD180*	2.2×10^{-6} (5.44)
		rs385839	19	5522356	-	5.5×10^{-6} (9.51)

		rs17839618	5	128477439	-	5.9×10^{-6} (4.88)
Chocolate preference	1585 (1493/92)	rs818534	5	32506309	*SUB1*	1.2×10^{-7} (4.26)
		rs17550268	17	33874678	*SLFN14*	2.4×10^{-6} (3.93)
		rs12944858	17	34067892	*RASL10B*	3.4×10^{-6} (2.97)
		rs7661128	4	156713031	*GUCY1B3*	5.4×10^{-6} (9.89)
Dark chocolate	1567 (1069/498)	rs969092	12	42234778	-	2.4×10^{-6} (3.68)
		rs1497177	12	42232982	-	3.2×10^{-6} (3.62)
		rs7460061	8	66377675	-	5.9×10^{-6} (1.44)

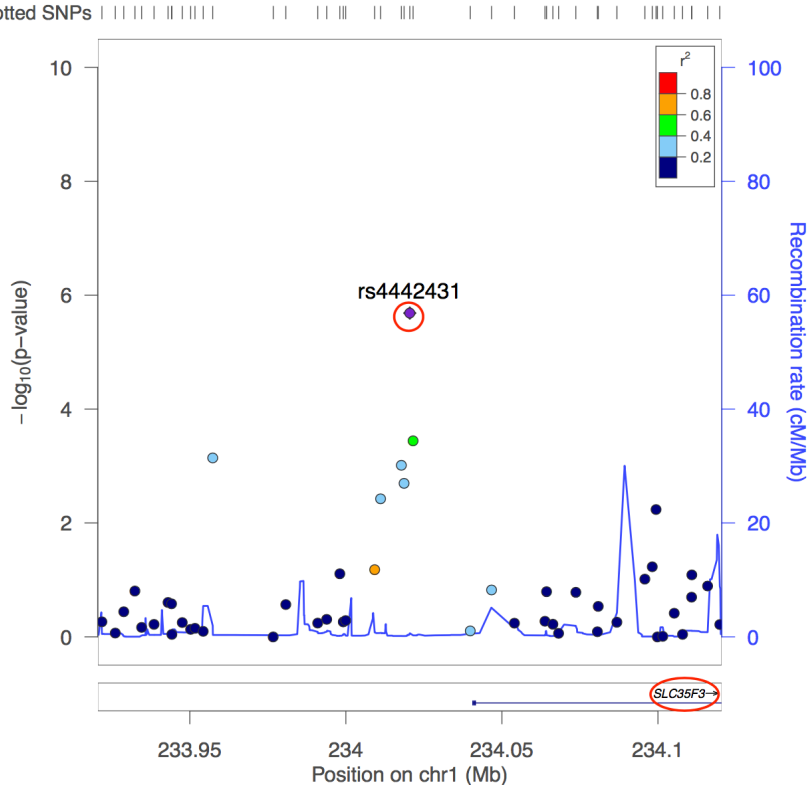
SNP, Single Nucleotide Polymorphism; Chr, Chromosome; [a]gene symbol from ensembl; [b]Fisher's exact test; OR, Odds ratio; *taste phenotype (bitterness and pepperiness for broccoli and watercress, respectively).

D.2 | Regional association plots for the GWAS of food preferences a) Butter and b) Broccoli. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 2 and 1. The purple-diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)

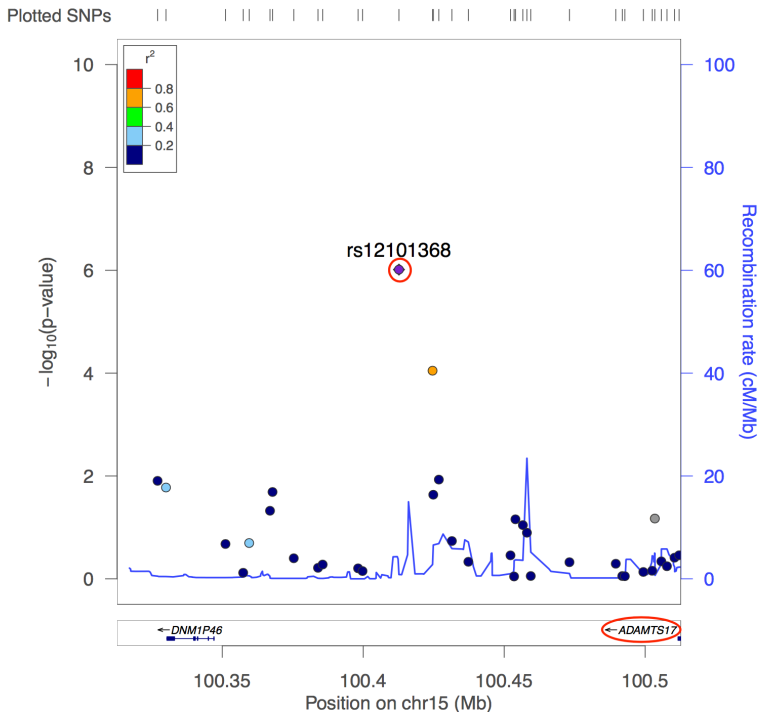


b)

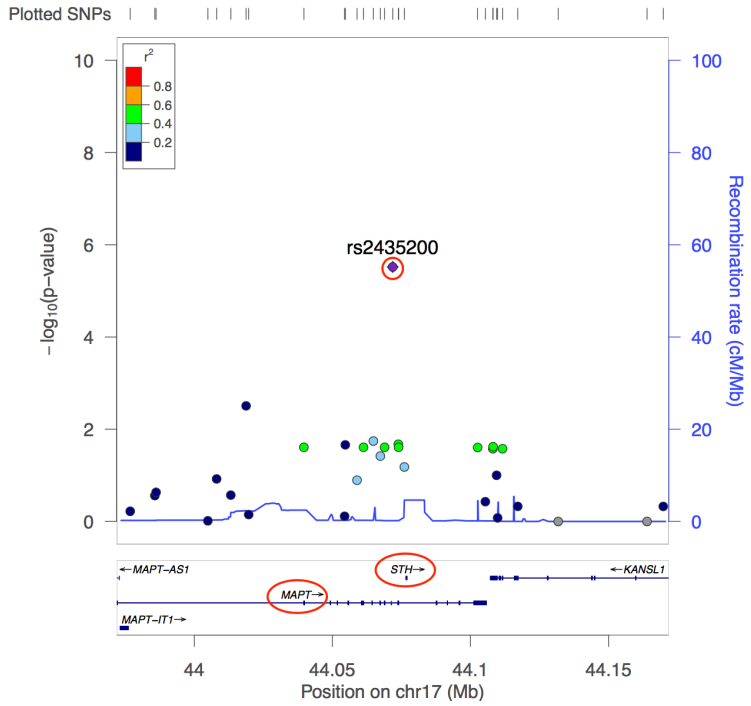


D.3 | Regional association plots for the GWAS of food preferences a) Soy sauce and b) Parmesan cheese. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 15 and 17. The purple-diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)

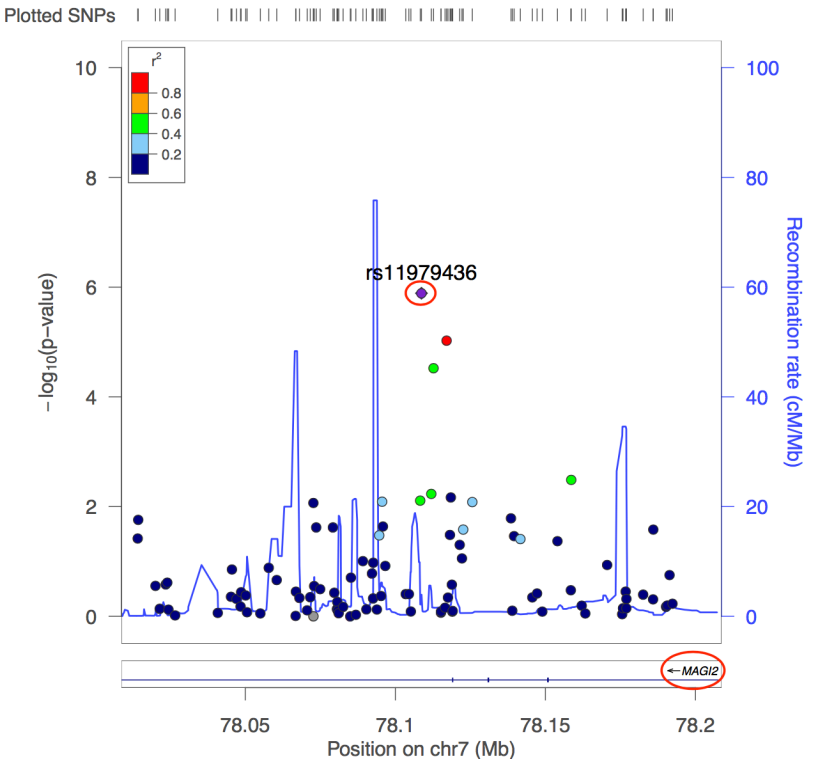


b)

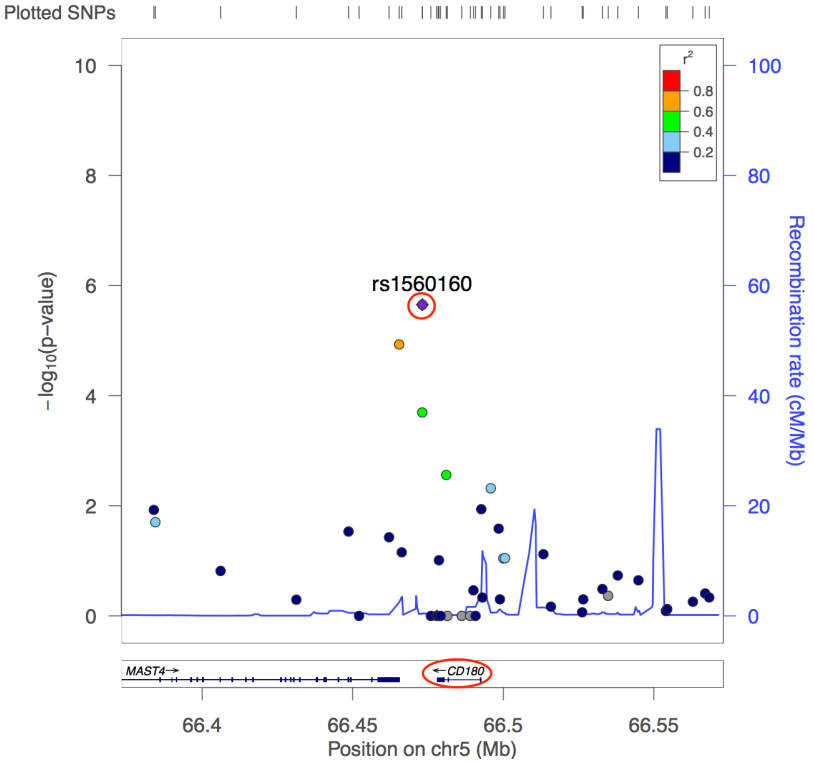


D.4 | Regional association plots for the GWAS of food preferences a) Marmite and b) Watercress. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 7 and 5. The purple-diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)



b)



D.5 | Distribution of rs1045643 genotypes in participants of Marmite preference GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Dislike	7	102	468	577
Like	14	215	509	738
Total	21	317	977	1315

$P_{Dom} = 4.8\times10^{-7}$, $OR_{Dom} = 1.93$; $P_{Rec} > 0.05$

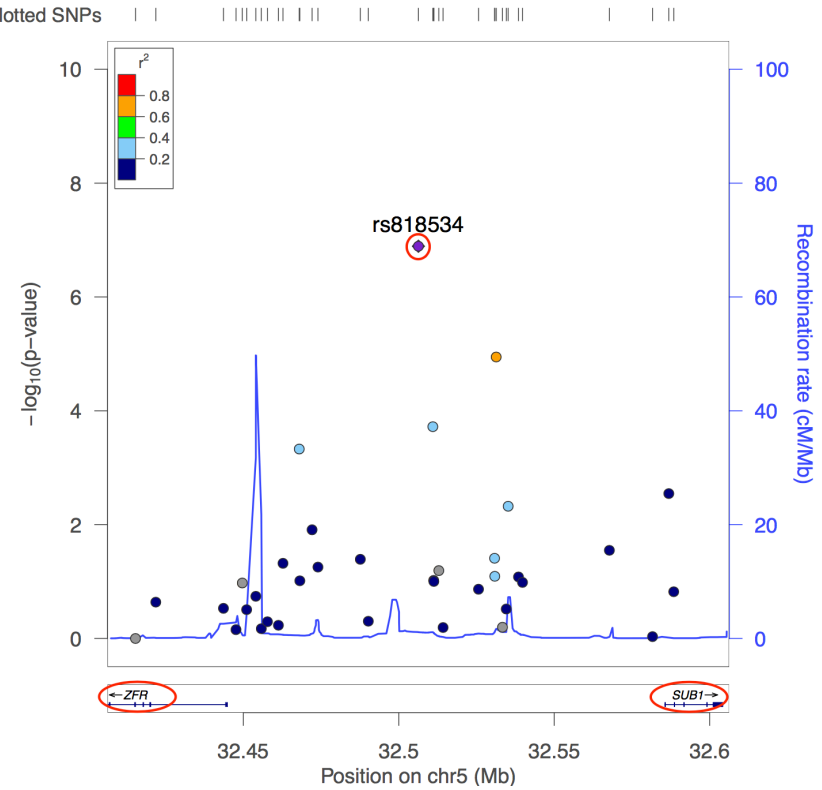
D.6 | Distribution of rs17550268 genotypes in participants of chocolate preference GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Dislike	1	12	79	92
Like	78	509	906	1493
Total	79	521	985	1585

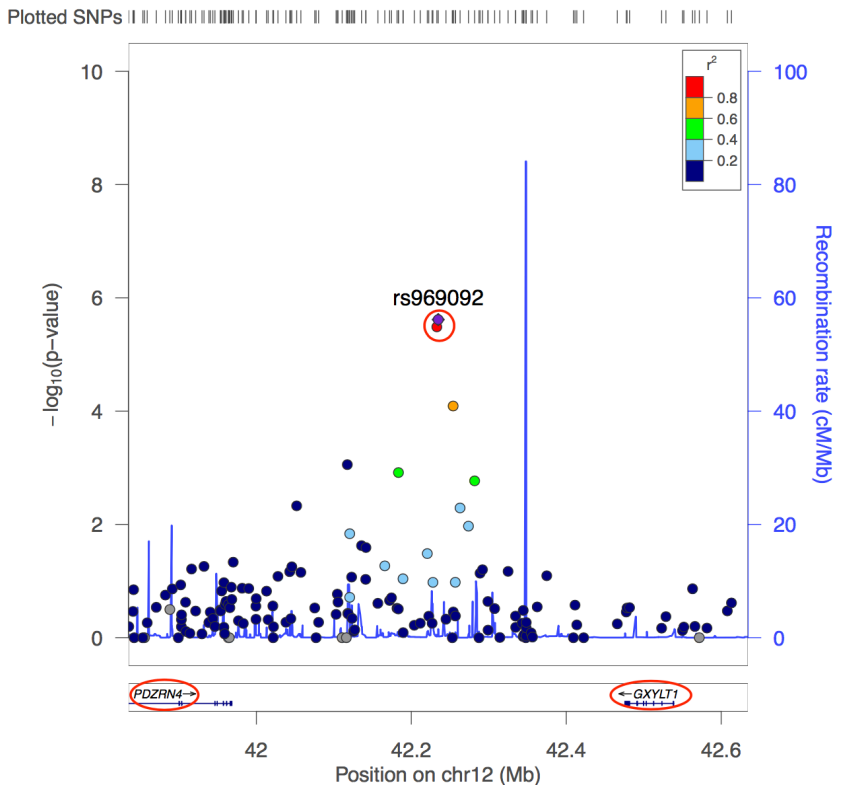
$P_{Dom} = 3.7\times10^{-7}$, $OR_{Dom} = 3.93$; $P_{Rec} = 0.08351$, $OR_{Rec} = 5.01$

D.7 | Regional association plots for the GWAS of food preferences a) Chocolate and b) Dark chocolate. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 5 and 12. The purple-diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)

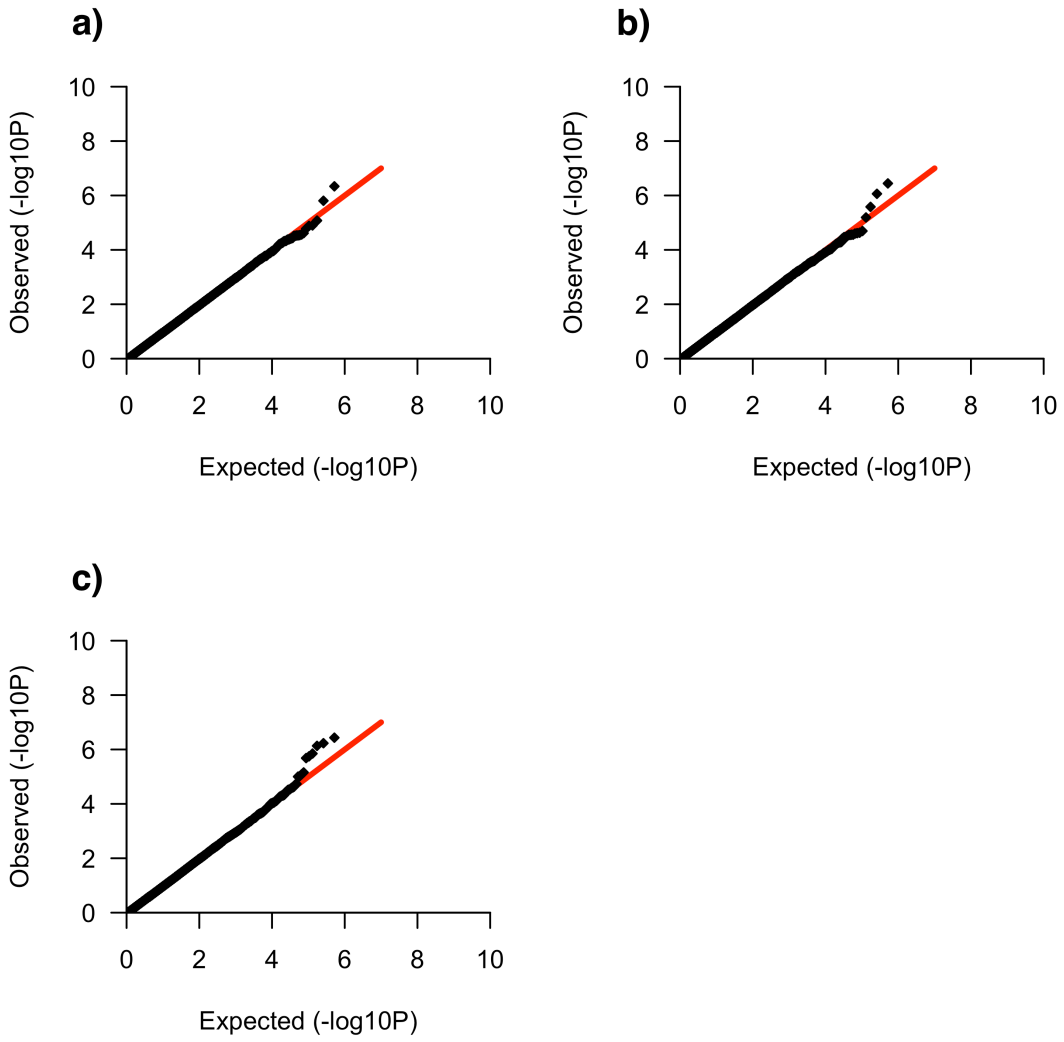


b)



Appendix E

E.1 | Q-Q plots for the GWAS of odour perception: a) Chocolate, b) Smoke and c) Banana. $-log_{10}$ of observed p-values (black dots) for each SNP were plotted against the $-log_{10}$ of expected p-values. The red line indicates the null hypothesis of no true association.



E.2 | Distribution of rs11632280 genotypes in participants of chocolate odour perception GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Not smelling	17	34	113	164
Smelling	69	483	653	1205
Total	86	517	766	1369

$P_{Dom} = 0.0004053$, $OR_{Dom} = 1.87$; $P_{Rec} = 0.03728$, $OR_{Rec} = 1.90$

E.3 | Distribution of rs227722 genotypes in participants of smoke odour perception GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Not smelling	18	31	102	151
Smelling	65	493	680	1238
Total	83	524	782	1389

$P_{Dom} = 0.003107$, $OR_{Dom} = 1.70$; $P_{Rec} = 0.002933$, $OR_{Rec} = 2.44$

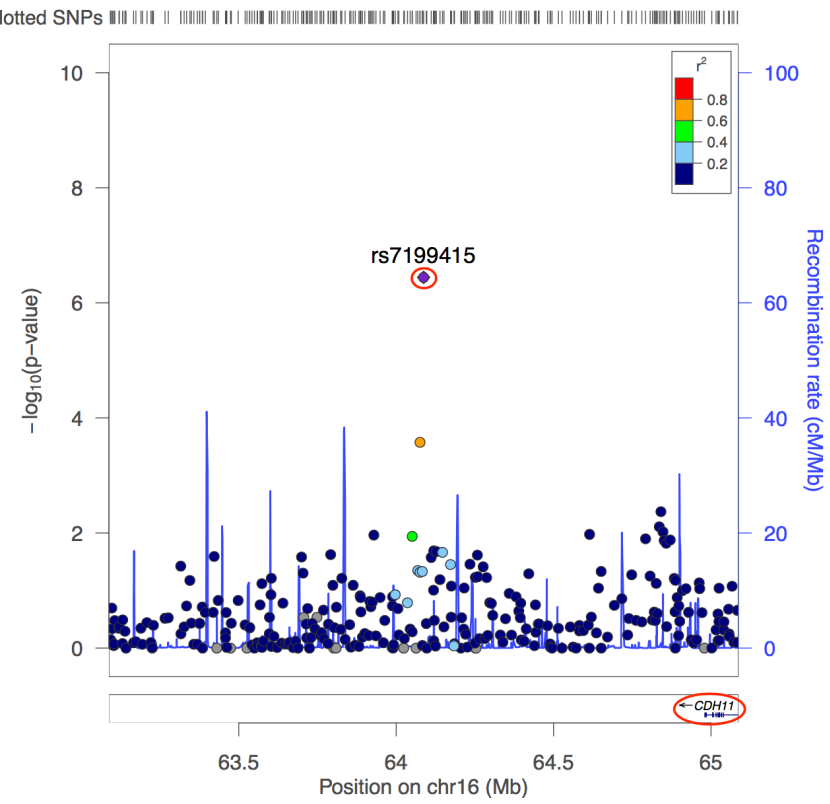
E.4 | Distribution of rs12543038 genotypes in participants of banana odour perception GWAS.

Genotype	***AA***	***Aa***	***aa***	***Total***
Not smelling	13	94	239	346
Smelling	8	179	789	976
Total	21	273	1028	1322

$P_{Dom} = 1.1 \times 10^{-5}$, $OR_{Dom} = 1.88$; $P_{Rec} = 0.0005429$, $OR_{Rec} = 4.71$

E.5 | Regional association plots for the GWAS of odour perception a) Smoke and b) Banana. Observed $-log_{10}$ p-values were plotted against the genomic locations on chromosome 16 and 10. The purple-diamond (circled in red) represents the strongest association. The colour-coded dots show the level of LD with the peak SNP. Genes are shown in the bottom.

a)



b)

