Exploring Associations between the Nicotinic Acetylcholine Receptor Gene Cluster *CHRNA5-A3-B4* and Smoking-Related Behaviours

Ph.D. Thesis

2012

Jennifer J. Ware

Summary

Tobacco use is the leading preventable cause of death worldwide. In order to address this epidemic, it is important that we have a thorough understanding of the aetiology of tobacco use and dependence. Twin and adoption studies have consistently demonstrated the importance of genetic factors in smoking behaviours. The advent of genome-wide technologies has greatly facilitated the search to determine which specific genetic factors contribute to tobacco use phenotypes. A locus within the nicotinic acetylcholine receptor gene cluster *CHRNA5-A3-B4* has generated particular interest – that marked by variants rs16969968 in *CHRNA5* and rs1051730 in *CHRNA3*. The primary aim of this thesis was to determine the role played by this locus in smoking-related behaviours, with an emphasis on phenotype refinement. A number of different approaches were utilised to address this objective, namely systematic review and meta-analysis, genetic epidemiology (including detailed phenotyping of smoking behaviour in adolescence), laboratory-based techniques, and genome-wide meta-analysis. Compelling evidence for a small, robust association was observed between the rs1051730/rs16966968 variants and daily cigarette consumption, equivalent to a per allele effect of approximately one cigarette per day. This effect was consistent across population sub-groups. Compelling evidence for an association between this locus and level of tobacco exposure was further illustrated through genome-wide meta-analysis of cotinine levels in current smokers. No association was observed between this locus and smoking initiation however, as examined in a prospectively assessed cohort using precisely defined phenotypes. An association between rs1051730/rs16969968 and smoking topography has yet to be explored. However, a full protocol was developed and piloted to investigate this. In addition, this research has also illustrated the importance of precise, objective, phenotype definition, an observation which has important implications for the fields of molecular genetics and epidemiology.

Statements and Declarations

Except where indicated by specific reference, the work submitted is the result of the candidate's own investigation and the views expressed are those of the candidate.

Candidate's signature: _____

No portion of the work presented has been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Candidate's signature: _____

Acknowledgements

First and foremost, I wish to thank my Ph.D. supervisors and mentors, Dr Marianne van den Bree and Prof Marcus Munafò. For the past four years they have walked a careful line between supporting me and challenging me, which has undoubtedly helped me to develop as a scientist. They have always found the time to answer my questions or discuss ideas, however trivial, and have taught me to be critical, questioning, and demanding of excellence. They have set the scientific standard I aspire to.

I also wish to extend my thanks to Dr Jon Heron from the University of Bristol, who was responsible for developing the smoking trajectories phenotype utilised in my genetic epidemiological study. He has also has been a fountain of statistical advice throughout my studies. In addition, I would like to thank Dr Nic Timpson, also from the University of Bristol, with whom the recall-by-genotype studies were developed.

I also wish to thank all those involved in the Cotinine Consortium (genome-wide meta-analysis project), including each contributing sample/cohort, alongside individual study leads and analysts responsible for conducting respective genome-wide association analyses which contributed to the meta-analysis. These individuals include Sam Chen (study lead for MESA, CARDIA and Framingham samples), Jaakko Kaprio (study lead for FinnTwin, GenMets and YFS), Tim Spector (study lead for TwinUK), Anu Loukola, Massimo Mangino, Alex Stiby, Leo-Pekka Lyytikainen, Juho Wedenoja and Kirsi Auro. Special thanks also go to Rachel Tyndale at the University of Toronto for her expert advice regarding determination of cotinine thresholds for this study.

Further gratitude is extended all those involved in Avon Longitudinal Study of Parents and Children, including all the families who took part in this study. This study has been an essential resource for the completion of my doctoral studies.

Special thanks go to the Wellcome Trust for their financial support throughout my studies, and to all those associated with the Wellcome Trust Ph.D programme,

including Miss Cath Hortop, Dr Vanessa Davies, and Prof Vincenzo Crunelli from Cardiff University.

From the University of Bristol, I wish to thank each and every member of the Tobacco and Alcohol Research Group, past and present. Thank you for keeping my spirits high and the debates lively.

Last but not least, I wish to thank my family, particularly my husband, Ryan, who has been my rock throughout the course of my studies.

Contents

List of Tables

List of Figures

Appendices

Abbreviations

ALSPAC	Avon Longitudinal Study of Parents and Children
CO	Carbon Monoxide
COPD	Chronic Obstructive Pulmonary Disease
CPD	Cigarettes per Day
DZ	Dizygotic
fMRI	Functional Magnetic Resonance Imaging
FTND	Fagerström Test for Nicotine Dependence
GWA	Genome-Wide Association
HWE	Hardy–Weinberg Equilibrium
IPN	Interpeduncular Nucleus
MAF	Minor Allele Frequency
MAO	Monoamine Oxidase
MHb	Medial Habenula
MZ	Monozygotic
ND	Nicotine Dependence
nAChRs	Nicotinic Acetylcholine Receptors
PANAS	Positive and Negative Affect Scale
QSU	Questionnaire of Smoking Urges
SNP	Single Nucleotide Polymorphism
UADT	Upper Aerodigestive Tract
WAIS-R	Wechsler Adult Intelligence Scale (Revised)

Chapter 1

Introduction

1.1 Tobacco use: The current picture

1.1.1 Prevalence and mortality

Tobacco use is one of the greatest public health concerns facing modern society. It currently accounts for the deaths of 5.4 million people a year – more than tuberculosis, HIV/AIDS and malaria combined (WHO, 2008). If current trends remain unchecked, it is estimated that tobacco use will account for the deaths of eight million individuals a year by 2030 (Mathers & Loncar, 2006). In order to address this growing epidemic, it is imperative that we have a thorough understanding of the aetiology of tobacco use and dependence.

> *"Tobacco is the only legally available consumer product which kills people when it is used entirely as intended"*
>
> (The Oxford Medical Companion, 1994, cited in WHO, 2008)

There are currently more than one billion smokers worldwide (WHO, 2008), approximately 10 million of which reside in Great Britain (ASH, 2012). In the UK, 20% of the adult population are current smokers (21% of adult males; 20% of adult females) (ONS, 2012), who consume an average of 13.1 cigarettes per day (NHS, 2011). Prevalence rates of smoking in the UK have been in decline since the peak noted in the late 1940s when official records began (65% adult males; 41% adult females) (NHS, 2011), although the rate of this decline has slowed dramatically in

recent years, with a mere 1% reduction in overall prevalence noted between 2007 and 2010 (ONS, 2012).

Whilst figures suggest that tobacco use is now falling in high-income countries (albeit gradually), the epidemic has shifted to the developing world where tobacco use is increasing (WHO, 2008). This is of particular concern given the substantial time delay between the peak in smoking prevalence and the subsequent peak in smoking-related mortality (see Lopez, Collishaw, & Piha, 1994).

1.1.2 Health consequences of tobacco use

Concerns regarding the potential health risks of tobacco use were raised as early as the late 18th century, namely relating use of tobacco pipes to cancer of the lip (Sommering, 1795, cited in Doll, 1998). Throughout the late 1920s and 1930s further evidence emerged relating smoking to cancers of the lip, mouth and lung (Lickint, 1929; Lombard & Doering, 1928; Muller, 1939), vascular disease (English, Willius, & Berkson, 1940) and decreased life expectancy (Pearl, 1938). In spite of this growing literature however, it wasn't until the publication of two case-control studies in the 1950s documenting an association between smoking and lung cancer (Doll & Hill, 1950; Wyndor & Graham, 1950) that the negative health impact of smoking finally began to gain general recognition. Subsequent prospective cohort studies served to reinforce these findings, of which Doll & Hill's classic study of the mortality of doctors in relation to their smoking habits remains a poignant example (Doll & Hill, 1954, 1956).

In 1962 the Royal College of Physicians (RCP) produced a report entitled 'Smoking and Health' which concluded that smoking was an important cause of lung cancer, and associated with a variety of other diseases including chronic bronchitis, pulmonary tuberculosis and coronary heart disease. The comprehensive U.S.

Surgeon General Report 'Smoking and Health' was published soon after in 1964 (U.S. Department of Health, Education and Welfare, 1964).

In the Surgeon General's 2004 report (U.S. Department of Health and Human Services, 2004), smoking was documented as a cause of cancers of the lung, larynx, oral cavity and pharynx, oesophagus, pancreas, bladder, kidney, cervix, stomach, and acute leukaemia. Smoking was also identified as a cause of multiple cardiovascular diseases (including coronary heart disease and stroke), numerous respiratory diseases (both acute (e.g., pneumonia) and chronic (e.g., chronic obstructive pulmonary disease)), fertility problems, and a broad spectrum of other diseases and disorders including loss of bone mass, dental diseases, erectile dysfunction and diseases of the eye.

Fifty years after the RCP publication of 'Smoking and Health', the list of diseases caused by smoking continues to expand. Smoking is now acknowledged to harm almost every organ in the body, and has been identified a risk factor for the six of the eight leading causes of death worldwide, namely ischemic heart disease, cerebro-vascular disease, lower respiratory infections, chronic obstructive pulmonary disease, tuberculosis and cancers of the trachea, bronchus and lung (WHO, 2008). The annual cost associated with treating smoking-related disease in the UK has been estimated at £5.2 billion (NHS, 2011).

1.2 Neurobiology of tobacco dependence and the addictive potency of tobacco

1.2.1 Introduction

In spite of a growing awareness of the negative health consequences of tobacco use, many continue to smoke. Whilst the majority of smokers express a desire to quit, they are unable to do so. A recent survey indicated that over 60% of smokers in the UK would like to stop smoking altogether (ONS, 2012). However,

only a minority of individuals (<5%) succeed in doing so without help in the long-term (Cohen et al., 1989; Hughes et al., 1992). Tobacco is highly addictive. Here I discuss the pharmacology and addictive properties of nicotine (the primary psychoactive drug present in tobacco), alongside other factors and design features of tobacco products that may facilitate or promote dependence.

1.2.2. Nicotine

Nicotine is the primary psychoactive drug present in tobacco. Although hotly disputed by the tobacco industry throughout the latter half of the 20th century (see Henningfield, Rose, & Zeller, 2006), it is now clear that nicotine is an addictive drug, characterised by compulsive use, psychoactive effects, and drug-reinforced behaviour, and has been recognised as such by the World Health Organisation (WHO) and the US National Institute on Drug Abuse, and in the 1988 report of the US Surgeon General (Henningfield et al., 2006).

> *"The strength and persistence of self-administration of a drug is perhaps the hallmark of its abuse liability, or ability to produce dependence"*
>
> (Donny, Caggiula, Knopf, & Brown, 1995, p390)

The addictive properties of nicotine have been extensively evidenced in both the human and animal literature (for an excellent review see Le Foll & Goldberg, 2009). Experimental paradigms for assessment include intravenous drug self-administration, conditioned place preference, drug discrimination, and measurement of withdrawal disturbances (Le Foll & Goldberg, 2009).

Nicotine, at relatively low doses, is a stimulant. It increases heart rate and blood pressure, and has beneficial effects on cognition and performance. A recent

meta-analysis has evidenced its positive effects on attention, memory, and fine motor skills (Heishman, Kleykamp, & Singleton, 2010). Tolerance to nicotine can develop rapidly (within a few days of use), and cessation of use results in withdrawal symptoms, both somatic and affective, such as anxiety, restlessness, inability to concentrate, irritability, and change in appetite (Stolerman & Jarvis, 1995).

Nicotine exerts its pharmacological effects through binding to nicotinic acetylcholine receptors (nAChRs). These receptors are widely distributed throughout the central and peripheral nervous system. They are ligand gated ion channels composed of 5 transmembrane subunit proteins arranged around a central pore (see Figure 1.1). Neuronal nAChRs consist of α (α_2-α_{10}) and β (β_2-β_4) subunits (Gotti, Zoli, & Clementi, 2006), each of which is encoded for by a single gene (denoted with a '*CHRN'* prefix), and may be homomeric or heteromeric in terms of subunit composition. Different combinations of subunits result in receptors differing in pharmacological and physiological profiles (Bierut, 2009; Paterson & Nordberg, 2000). Individual subtypes differ, for example, in their affinity for nicotine, and sensitivity to upregulation and desensitisation following nicotine exposure (Paterson & Nordberg, 2000). Each receptor subtype has a distinct distribution profilewithin the brain (see Figure 1.2). The distribution of specific receptor subtypes within the brain has been determined through assessment of subunit mRNA using techniques such as in situ hybridisation, and also through imaging techniques such as PET and SPECT, using subtype selective radioligands (Paterson & Nordberg, 2000). The differential expression of specific subunits with distinct biological functions in brain regions mediating specific behaviours allows nicotine to exert a broad range of effects (Decker, Sullivan, Arneric, & Williams, 2000). The $\alpha_4\beta_2$ receptor subtype is the most commonly expressed subtype in the human brain, and has historically been

implicated by animal models as critical to the experience of nicotine's reinforcing effects (e.g., Picciotto et al., 1998). In recent years however, the importance of the role played by the lesser studied α_3 and α_5 receptor subunits in nicotine dependence has been recognised, as is discussed later at length. For a detailed description of the regional distribution of these subunits see Improgo, Scofield, Tapper, & Gardner (2010).

Figure 1.1. Structure of nicotinic acetylcholine receptors (figure reproduced from Changeux, 2010).

[This image has been removed by the author for copyright reasons]

Figure 1.2. Regional distribution of nicotinic acetylcholine receptors in the rodent central nervous system (figure reproduced from Gotti et al., 2006).

[This image has been removed by the author for copyright reasons]

Nicotine exerts its complex effects (arousal, mood modulation, pleasure) via several neurotransmitter pathways. Once bound to neuronal nAChRs, it facilitates the release of dopamine, serotonin, and a host of other neurotransmitters including GABA, glutamate, norepinephrine, acetylcholine and endorphins (Benowitz, 2008). The mesolimbic dopamine pathway has perhaps been the most widely studied in relation to nicotine dependence (Balfour, 2002). Dopamine release in the nucleus accumbens, resulting from nicotinic stimulation of dopaminergic neurons in the ventral tegmental area, is crucial to the processing of rewarding and reinforcing effects of nicotine. Indeed, dopamine release in the nucleus accumbens appears critical in the experience of rewarding effects of many drugs of abuse.

Chronic exposure to nicotine results in a number of neuroadaptions (Balfour, 2002). These include desensitisation of nAChRs, alongside an upregulation in their expression (Benowitz, 2008), factors linked to nicotine tolerance and withdrawal.

Continued pairing of the rewarding/reinforcing effects of nicotine with specific sensory and environmental stimuli (such as the smell of tobacco smoke, or the sight of a pack of cigarettes) results in these stimuli acquiring reinforcing properties. These cues (conditioned reinforcers) have been linked to the maintenance of smoking, smoking-related cravings and relapse (Benowitz, 2008).

1.2.3 Additional constituents and design features promoting dependence

Whilst nicotine is the key psychoactive drug found in tobacco, the addictive potency of cigarettes (and indeed other tobacco products) is likely influenced by product design and inclusion of a number of additives. Indeed, the addictive potency of cigarettes is higher than that of pure nicotine products (Henningfield & Zeller, 2002). The modern cigarette is a sophisticated drug delivery device, carefully

tailored/engineered to optimise delivery of nicotine (and other constituents) to the smoker. The release of previously secret industry documents has evidenced this (see Hurt & Robertson, 1998). In this section we shall consider the role of tobacco smoke constituents, additives and design features of cigarettes that may facilitate/sustain tobacco dependence.

Sugars and polysaccharides are naturally present in tobacco, and commonly added to tobacco products in substantial quantities (SCENIHR, 2010). By themselves, these additives are not addictive, however, when burned they form numerous aldehydes e.g., formaldehyde and acetaldehyde. Acetaldehyde has not only been shown to have addictive potential in and of itself, as demonstrated through self-administration experiments in animals (Philip Morris, 1992), but has also been shown to enhance the addictive potential of nicotine. Moreover, the interaction between these compounds was observed to result in a rewarding effect that exceeded the additive effects of both in rodent studies (Philip Morris, 1992). Study of the interactions between nicotine and other smoke constituents is crucial to our understanding of the addictiveness of cigarettes.

Monoamine oxidase (MAO) inhibitors are also present in tobacco smoke. MAO inhibitors increase levels of certain amines in the brain, such as dopamine and serotonin, and may subsequently potentiate the reinforcing effects of nicotine (Hatsukami et al., 2010). Indeed, animal studies have demonstrated that MAO inhibitors facilitate nicotine self-administration and enhance its motivational properties (Guillem et al., 2005; Villegier et al., 2006). Such findings may, in part, explain why cigarettes have much stronger reinforcing properties than pure nicotine.

Menthol and other flavourings (e.g., clove, liquorice) not only serve to increase the palatability of cigarette smoke, and, in the case of menthol and clove,

facilitate deeper inhalation and therefore higher nicotine dose (due to their cooling/local anaesthetic effects), but may also become conditioned reinforcers in themselves as a consequence of their repeated pairing with nicotine (Carter et al., 2009). In addition, menthol also inhibits metabolism of nicotine to cotinine, purportedly through inhibition of CYP2A6 enzyme activity (Benowitz, Herrera, & Jacob, 2004), thus increasing the effect of nicotine. Cocoa and chocolate, which contain theobromine, are also common additives in tobacco. Theobromine is a bronchodilator, and thus has been proposed to enhance nicotine absorption in the lungs. However, theobromine content of cigarettes was deemed too low to exert bronchodilation in a recent review (SCENIHR, 2010).

Alkaline additives such as ammonia compounds are among the most abundant additives used in cigarette manufacture (Hurt & Robertson, 1998), added to cigarettes (and other tobacco products) in order to manipulate pH. Nicotine exists in both bound and unbound ("free") forms, dependent on pH. Unbound forms, abundant in alkaline conditions, are more physiologically active than bound forms, able to cross biological membranes into the bloodstream with greater ease. Industry scientists have extensively investigated the potential of pH manipulation to optimise nicotine delivery, and have fully exploited methods to increase tobacco smoke pH in order to maximise nicotine "kick" (see Hurt & Robertson, 1998), and, perhaps, to determine region of nicotine absorption: It is of note that the high buffering capacity of the lung lining fluid may limit the impact of smoke pH on nicotine absorption in the lungs (SCENIHR, 2010). In contrast, absorption of nicotine across the oral mucosa is more dependent upon pH (of particular relevance to cigars).

A number of physical characteristics of cigarettes have been engineered to manipulate nicotine delivery, including cigarette dimensions, filtration, ventilation,

paper porosity, and tobacco shred size (Hurt & Robertson, 1998). Ventilation, for example, serves to manipulate nicotine, tar and carbon monoxide levels through dilution of tobacco smoke, and is achieved through the introduction of holes in both the filter and paper wrap (SCENIHR, 2010). Ventilation technology was utilised in the production of "light" or "low-tar" cigarettes, which were promoted by the tobacco industry as healthier alternatives to full-strength cigarettes. Such labels have proved misleading however. Whilst smoking machine assessments suggest that these cigarettes yield lower doses of nicotine, studies have demonstrated that smokers compensate (e.g., through deeper inhalation, increased number of puffs per cigarette and so on) when smoking these cigarettes in order to achieve the same dose of nicotine attained whilst smoking stronger brands (Strasser, Lerman, Sanborn, Pickworth, & Feldman, 2007). Ventilation is also purported to effect particle size of tobacco smoke aerosol, which may impact on nicotine absorption into the bloodstream, although evidence for this is unclear.

1.3 Genetics and smoking behaviour

1.3.1 Heritability

Heritability is a measure of the degree of phenotype variability in a population that is attributable to genetic variation. Twin and adoption studies have allowed us to determine the relative influences of genetic and environmental factors on smoking-related behaviours. These approaches are discussed in detail below.

Adoption studies: Adoption studies have allowed us to disentangle the influence of genetic and environmental factors as causes of family resemblance (Plomin, DeFries, McClearn, & McGuffin, 2008). Examination of the degree of resemblance between adopted individuals and their biological parents enables us to determine genetic effects, whilst examination of the resemblance between adoptees

and their adoptive parents allows us to determine familial environmental influences (Sullivan & Kendler, 1999).

Twin studies: Classical twin study designs have been used since the early 1920s to estimate genetic influence on behaviours/traits (hereto referred to as phenotypes) (see Merriman, 1924). Monozygotic (MZ) twins are considered to be genetically identical (with minor exceptions such as individual de novo somatic mutations). In contrast, dizygotic (DZ) or "fraternal" twins share approximately 50% genetic similarity, as observed between full siblings. Presumably, all twin pairs growing up together share the same common environment. Comparison of the correlation of phenotypes within MZ twin pairs with those of DZ twin pairs allows us to determine the degree to which that phenotype is under genetic and environmental influence. Statistical modelling may be used to determine variation in liability to a phenotype attributable to separate genetic and environmental components.

Twin and adoption studies have provided consistent evidence that genetic factors contribute to the aetiology of cigarette smoking, playing an important role in smoking initiation, progression to heavy use and persistence (Fowler et al., 2007; Kendler et al., 1999; Lessov et al., 2004; Munafo & Johnstone, 2008; Sullivan & Kendler, 1999). A degree of variation in heritability estimates has been observed, as is to be expected given different time periods of assessment and different populations studied (Kaprio, 2009). For reference however, a recent meta-analysis (Li, Cheng, Ma, & Swan, 2003) reported that genetic factors were responsible for approximately 50% of the variation noted in smoking initiation, and approximately 60% of variation in smoking persistence.

As an aside, it is important to bear in mind that genetic and environmental factors do not always act as independent factors. Environmental factors can, for example, influence expression of genetic effects. Whilst the study of gene × environment (G×E) interactions in the context of behavioural phenotypes has proved controversial (Flint & Munafo, 2008; Riley, 2008; Uher, 2008), smoking is one case where there is *a priori* evidence of interaction – whatever one's genetic risk, it is not possible to become tobacco dependent without first exposing oneself to tobacco. Environmental factors may plausibly influence the probability of tobacco experimentation, and thereby moderate the expression of genetic liability for subsequent dependence. These issues are discussed at length in Chapter 3.

1.3.2 Molecular genetics

Genetic epidemiological studies, primarily exploiting classical twin designs, have provided a wealth of evidence demonstrating the importance of genetic factors as a whole in the aetiology of smoking-related behaviours. Advances in the identification of *specific* genetic variants associated with such phenotypes are now being made in the field of molecular genetics. Multiple approaches have been utilised to identify specific genes and their relationship to specific smoking-related phenotypes, namely linkage analysis, candidate gene association studies and, more recently, genome-wide association (GWA) studies. These approaches are discussed below.

> *"In genetic epidemiology, gene variation is not measured directly. Instead, the action of genetic and environmental factors is inferred from patterns of resemblance in special classes of relatives, particularly twins and adoptees. Molecular genetic studies relate disease risk directly to DNA variation"*
>
> (Kendler et al., 2012, p.181).

Linkage studies. Linkage studies employ a within-family design. This approach has traditionally been employed to study single-gene disorders, but may also be utilised to identify chromosomal regions associated with complex diseases such as smoking, for example through the study of sibling pairs (Dawn Teare & Barrett, 2005). Linkage studies of smoking-related phenotypes have identified a number of chromosomal regions, although regions identified using this approach have proven largely inconsistent across studies, and findings have not mapped well onto those stemming from candidate gene studies (David & Munafo, 2008).

Candidate gene association studies. This approach involves the comparison of allele frequencies between two groups of individuals at a pre-specified genetic locus. These groups are selected on the basis of a specific phenotype, and typically comprise a 'case' and a 'control' group (e.g., ever smokers versus never smokers), or groups displaying extremes of a trait (e.g., heavy smokers versus light smokers). As a specific genetic variant has to be selected for study, this approach is hypothesis driven (in contrast to linkage and genome-wide association studies). Variants selected for this approach are of known function, or linked to variants of known function, and are selected for the study of a specific phenotype on this basis, hence the name 'candidate gene' study. Candidate gene studies of smoking behaviour have focused primarily on targets within relevant neurotransmitter pathways (e.g., dopamine pathway genes) and enzymes associated with nicotine metabolism (e.g., CYP2A6) (Munafo & Johnstone, 2008). This approach has certain limitations. Firstly, as highlighted, *a priori* hypotheses are required with regards to candidature. Secondly, given that effects of individual variants in complex diseases tend to be very small, very large sample sizes are required to detect them. This may underlie persistent failures to replicate.

Genome-wide association studies. Genome-wide association studies adopt the same approach to sample selection and analysis as candidate gene studies (i.e., comparison of allelic frequencies between groups selected on the basis of phenotype). However, instead of focusing on a specific (or handful of) pre-specified locus (loci), microarrays ('gene chips') are used to systematically genotype hundreds of thousands of genetic polymorphisms across the genome. As such, and in contrast to candidate gene studies, they are strictly agnostic in their approach, requiring no *a priori* hypothesis. The advent of genome-wide technologies has greatly facilitated the search to determine which genetic variants contribute to specific diseases, including smoking behaviours (Furberg et al., 2010; Liu et al., 2010; Thorgeirsson et al., 2010), and have aided identification of variants which would perhaps not have been considered previously on the basis of biological function.

Whilst molecular genetics has certainly advanced our knowledge of the genetic underpinnings of smoking behaviour, we are still far from a full and comprehensive understanding. This area has been notoriously hampered by failures to replicate promising initial findings. One major exception to this rule however concerns a locus within the nicotinic receptor gene cluster *CHRNA5-A3-B4*. This is discussed at length in the following section.

1.4 The nicotinic acetylcholine receptor gene cluster *CHRNA5-A3-B4*

1.4.1 Background

Nicotinic acetylcholine receptors (nAChRs), to which nicotine binds, serve as the 'gateways' through which nicotine exerts its effects on the brain, as previously discussed. Recent years have witnessed a rapid growth in research focused on the nAChR gene cluster *CHRNA5-A3-B4* on the long arm of chromosome 15 (15q24-25.1), responsible for encoding three nAChR subunits (α_5, α_3, and β_4). One locus

within this cluster has generated particular interest – that marked by the single nucleotide polymorphisms (SNPs) rs16969968 in *CHRNA5* and rs1051730 in *CHRNA3*. These highly correlated SNPs**,** which at the time of writing have been broadly studied, are now firmly established predictors of multiple smoking-related behaviours and diseases**,** and form the focus of this thesis. Within this section we discuss the initial discovery of an association between these variants and smoking behaviour, the numerous phenotypes with which they have been subsequently associated (smoking-related behaviours, diseases, and cognitive phenotypes; see Tables 1.1, 1.2 and 1.3 respectively), and potential mechanisms purported to underlie such associations. Gene × environment interactions are also discussed, alongside issues relating to phenotype definition and measurement precision.

Table 1.1. Smoking behaviours associated with rs1051730/rs16969968

Author	Year	Original research or meta-analysis	Phenotype/s associated with SNP/s	Notes
Xie	2011	Original	ND	Interaction noted between childhood adversity and rs16969968 genotype in predicting ND in males only (no interaction seen in females). No main effect of rs16969968 noted however.
Breetvelt	2011	Original	Smoking quantity	Association noted between rs16969968 and CPD and heavy smoking (25+ cig/day). No association noted with cessation, lifetime smoking, or current smoking.
Munafo	2011	Original	Smoking cessation	Weak evidence of an association between rs1051730 and short-term smoking cessation. No evidence of association at later follow-up.
Timofeeva	2011	Original	Cotinine; Lung cancer risk	rs16969968 associated with circulating cotinine levels and lung cancer risk.
Siedlinski	2011	Original	Smoking quantity	Nominally significant association noted between rs1051730 and lifetime average CPD, but not current CPD. Sample consisted of 4 cohorts of ever smokers with COPD.
Marques-Vidal	2011	Original	Smoking quantity; ND; Difficulty quitting (borderline)	rs1051730 associated with heaviness of smoking (assessed using heaviness of smoking index), ND, and nominally associated with difficulty quitting (although this effect was no longer apparent after adjusting for nicotine dependence). No association found with willingness/attempt/preparation to quit.
Wassenaar	2011	Original	Smoking quantity; ND; Lung cancer risk (*sig. after adjustment for pack-years*)	rs1051730 associated with CPD, ND (FTND score), and lung cancer risk.
Kaur-Knudsen	2011	Original	Smoking quantity; Pack-years; Lung cancer; Bladder cancer; COPD.	rs1051730 associated with lung cancer, bladder cancer and COPD after adjustment for smoking. No association with ischemic heart disease or ischemic stroke. rs1051730 was associated with smoking quantity (g/day) and pack-years (cumulative tobacco consumption), but not with smoking status, age of initiation, age of cessation, or smoking duration. Very large sample (>10k).
Hong	2011	Original	Smoking severity; Schizophrenia	rs16969968 associated with smoking severity (measured using FTND) in both smokers with schizophrenia and control smokers. This variant was not associated with smoking status however in either schizophrenia patients or controls.

Rodriguez	2011	Original	Continued smoking following experimentation	rs16969968 genotype, acting recessively, affects smoking (namely continued smoking in those who experiment) similarly in adolescents (13-15 years) and adults (18 years). Examination of ORs suggests slightly larger effect at 18 years. No association noted between rs16969968 and 'experimentation' (assessed as current & past vs. never smokers) at either age.
Lori	2011	Original	ND; Cotinine level	rs16969968 nominally associated with ND (FTND score) and cotinine level.
Sorice	2011	Original	Smoking quantity	rs1051730 associated with smoking quantity in two of three Italian populations (differing in environment, history, and genetic structure).
Sarginson	2011	Original	Smoking quantity	Pharmacogenetic study. rs16969968 and rs1051730 associated with baseline smoking quantity. No association noted between either SNP and baseline ND as assessed using the modified Fagerstrom Test of Nicotine Dependence (mFTQ). Neither SNP associated with abstinence, craving or withdrawal during treatment, although other SNPs in the 15q24 region were.
Ducci	2011	Original	Heavy/regular smoking (in adolescence and adulthood)	rs1051730 associated with heavy/regular smoking (non-smoker control groups), with similar effect of SNP noted at age 14 and 31 years. No association noted between rs1051730 and occasional/light smoking (non-smoker control groups) however at either age, suggesting that this SNP is not involved in initiation.
Kim	2011	Original	Emphysema (severity); ND	rs1051730 (and rs8034191) associated with ND (FTND assessed). Both SNPs also associated with severity of emphysema, but only in *former* smokers, not current smokers.
Winterer	2010	Original	ND; Cognition	rs1051730 and rs16969968 associated with ND and cognitive performance (cognitive domains from the WAIS-R, and n-back task performance).
Johnson	2010	Original	ND	rs16969968 associated with ND. An interaction between rs16969968 genotype and peer smoking was noted in predicting ND risk.
Saccone	2010	Meta-analysis	Lung cancer; COPD (nominal); Smoking quantity	rs16969968 associated with lung cancer and COPD (marginal), *after adjusting for CPD*. rs16969968 associated with CPD.
De Ruyck	2010	Original	ND (borderline)	rs1051730 marginally associated with FTND score. No association noted between this SNP and smoking cessation (abstinence assessed 1 week, 1 month and 6 months after short-term nicotine patch treatment) or withdrawal symptoms however.
Grucza	2010	Original	ND	rs16969968 associated with ND. rs16969968 exhibited a larger effect in later-onset (post 16 years) smokers (contrasts Weiss et al., 2008).

Thorgeirsson	2010	Meta-analysis	Smoking quantity	rs1051730 associated with CPD.
Furberg	2010	Meta-analysis	Smoking quantity	rs1051730 associated with CPD.
Liu	2010	Meta-analysis	Smoking quantity	rs1051730 associated with CPD.
Lips	2010	Original	Smoking quantity; Lung cancer; UADT cancer	rs16969968 associated with CPD and heavy smoking. An association was also noted with lung cancer (effect virtually unchanged following adjustment for smoking), and earlier age of lung cancer onset. Association also noted with UADT cancers. NO association with smoking initiation or cessation.
Chen, Johnson	2009	Original	ND	rs16969968 associated with ND. Risk for ND associated with rs16969968 modified by level of parental monitoring – risk increased significantly with the risk genotype of this SNP when combined with lowest quartile parental monitoring.
Keskitalo	2009	Original	Smoking quantity; Cotinine	rs1051730 associated with both CPD and serum cotinine level. Notably, proportion of variance accounted for by rs1051730 was five times greater for cotinine relative to CPD (4.3% vs. 0.9%).
Chen, Chen	2009	Original	ND; Symptoms of alcohol abuse	rs1051730 and rs16969968 associated with FTND score. Both SNPs were also associated with symptoms of alcohol abuse/dependence, but the associated alleles were the opposite of that of FTND. NO association noted with cannabis abuse/dependence.
Saccone	2009	Original	ND	rs16969968 associated with ND in full sample, and in separate European American and African American subsamples.
Freathy	2009	Original	Smoking cessation; Smoking quantity	rs1051730 associated with smoking quantity and reduced ability of women to quit smoking during pregnancy.
Saccone	2009	Original	ND	rs16969968 and rs1051730 associated with ND.
Caporaso	2009	Original	Smoking quantity	rs1051730 associated with CPD.
Breitling	2009	Original		Neither rs1051730 nor rs16969968 associated with cessation in ever heavy (>20 CPD) smokers.
Weiss	2008	Original	ND severity (dependent on age of initiation)	rs16969968 and rs1051730 associated with severity of ND among long-term smokers who began daily smoking before age 16, but not among those who began daily smoking post 16 years.

Le Marchand	2008	Original	Urinary concentration of nicotine equivalents; Carcinogenic tobacco-specific nitrosamine levels	rs1051730 and rs16969968 are associated with smoke exposure, as determined using measures assessing urinary concentrations of nicotine and its metabolites. This association survives adjustments for CPD, leading authors to conclude that "simple adjustment for number of cigarettes per day is inadequate to control for smoking dose in studies examining the independent association of these variants with smoking-associated lung cancer".
Sherva	2008	Original	Smoking status; First smoking experiences.	rs16969968 associated with smoking status (i.e., ever [≥5 CPD for ≥ 5 years] vs. never [<100 cigs consumed in lifetime]), and, in Caucasians, experiencing a pleasurable buzz during the first cigarette.
Spitz	2008	Original	CPD; FTND; age at onset of lung cancer	No evidence of association between rs1051730 and lung cancer in never smokers. Partial sample overlap with Amos et al. Age at onset of lung cancer was modified by genotype (risk genotype = earlier age of onset). Highest risk for variant noted in *lightest* smokers. No association between rs1051730 and bladder/renal cancer.
Thorgeirsson	2008	Original	Smoking quantity; ND; Lung cancer; Peripheral arterial disease	rs1051730 associated with CPD, ND, lung cancer and peripheral arterial disease.
Bierut	2008	Original	Habitual smoking; nAChR function	rs16969968 and rs1051730 associated with habitual smoking (case = 20+ CPD for 6mths+; control = 100+ cigs consumed in lifetime or had smoked daily for 1mth+ but never consumed >10CPD). rs16969968 associated with nAChR A5 subunit function (minor allele results in subunit less responsive to nicotine agonist) but not expression.
Saccone	2007	Original	ND	rs16969968 associated with ND in candidate gene study.

CPD = cigarettes per day, ND = nicotine dependence; FTND = Fagerström Test for Nicotine Dependence; COPD = chronic obstructive pulmonary disease; SNP = single nucleotide polymorphism; UADT = upper aerodigestive tract; WAIS-R = Wechsler Adult Intelligence Scale (Revised).

Table 1.2. Diseases associated with rs1051730/rs16969968

Author	Year	Original research or meta-analysis	Phenotype/s associated with SNP/s	Notes
Timofeeva	2011	Original	Cotinine; Lung cancer risk	rs16969968 associated with circulating cotinine levels and lung cancer risk.
Jaworowska	2011	Original	Lung cancer	rs16969968 associated with lung cancer risk, but not bladder or laryngeal cancer, in Polish population.
Wassenaar	2011	Original	Smoking quantity; ND; Lung cancer risk (*sig. after adjustment for pack-years*)	rs1051730 associated with CPD, ND (FTND score), and lung cancer risk.
Xun	2011	Original		rs16969968 was not associated with survival time in a large cohort of lung cancer patients (regardless of whether cause of death was from lung cancer or not). Stratified analyses suggested a role for rs16969968 in influencing survival time in *never*-smoking lung cancer patients (all-cause mortality, and lung-cancer specific mortality to a lesser extent) although this is possibly due to small sample of never smokers. Study based on cohort from Hung et al.
Chen, Wu	2011	Original		rs1051730 was not associated with pancreatic cancer risk. NB: Controls matched for smoking behaviour.
Kaur-Knudsen	2011	Original	Smoking quantity; Pack-years; Lung cancer; Bladder cancer; COPD.	rs1051730 associated with lung cancer, bladder cancer and COPD after adjustment for smoking. No association with ischemic heart disease or ischemic stroke. rs1051730 was associated with smoking quantity (g/day) and pack-years (cumulative tobacco consumption), but not with smoking status, age of initiation, age of cessation, or smoking duration. Very large sample (>10k).
Hong	2011	Original	Smoking severity; Schizophrenia	rs16969968 associated with schizophrenia in both Caucasian and African-American non-smoker schizophrenia patients compared with control non-smokers. This variant was not associated with smoking status however in either patients or controls.
Chen, Gorlov	2011	Original	Tumor size at diagnosis	rs1051730 associated with larger tumor size at diagnosis (squamous cell carcinoma).
Chen, Truong	2011	Original	UADT cancers	rs16969968 associated with UADT cancers in women (but not men). No evidence for a sex effect on relationship between rs16969968 and CPD.

Sakoda	2011	Original	Lung cancer	rs16969968 associated with lung cancer risk. Diet had little impact on this relationship. A stronger association was noted however in individuals diagnosed at <70yrs of age, and in those with a baseline smoking history of <40 pack/years.
Gago-Dominguez	2011	Original	Bladder cancer (borderline)	Association noted between rs8034191 (highly correlated with rs1051730) and bladder cancer which persisted after adjustments for CPD and number of years smoking. Borderline association noted with rs1051730.
Kim	2011	Original	Emphysema (severity); ND	rs1051730 and rs8034191 associated with severity of emphysema, but only in *former* smokers, not current smokers.
Erlich	2010	Original	Opioid dependence severity	rs16969968 associated with opioid dependence severity.
Amos	2010	Original	Lung cancer	rs1051730 and rs16969968 associated with lung cancer in African American sample.
Saccone	2010	Meta-analysis	Lung cancer; COPD (nominal); Smoking quantity	rs16969968 associated with lung cancer and COPD (marginal), *after adjusting for CPD*. rs16969968 associated with CPD.
Wang	2010	Original	Lung cancer; COPD	Evaluated the role of smoking behaviour (pack-years) and COPD (both alone and in combination) as mediators of the relationship between rs1051730 and lung cancer. Also examined the mediating effect of smoking behaviour on the relationship between rs1051730 and COPD. Concluded that rs1051730 is both directly and indirectly associated with lung cancer. Pack-years was shown to be a mediator, yet COPD was "a more significant mediator than pack years" (11.5% vs. 7.6%). rs1051730 was also associated with COPD (pack-years also shown to mediate this relationship).
Truong	2010	Original	Lung cancer	rs16969968 associated with lung cancer in white ever smokers. This association was observed for all histology types (adenocarcinoma/squamous/large cell/small cell), and was stronger for those diagnosed at younger ages. No association noted in never smokers or Asians. Very large sample (12k cases, 15k controls).
Hansen	2010	Original	Lung cancer	African American (AA) sample. A four SNP haplotype spanning CHRNA5 (including rs16969968) and CHRNA3 was associated with increased lung cancer risk. rs16969968 was not singularly associated with lung cancer (possibly due to low frequency of risk variant in AAs). rs1051730 was not associated with lung cancer risk.
Lips	2010	Original	Smoking quantity; Lung cancer; UADT cancer	rs16969968 associated with lung cancer (effect virtually unchanged following adjustment for smoking), and earlier age of lung cancer onset. Association also noted with UADT

				cancers. NO association with smoking initiation or cessation.
Lambrechts	2010	Original	Emphysema	rs1051730 genotype associated with the presence and severity of emphysema. This association was independent of pack-years smoking.
Girard	2010	Original		rs1051730 was NOT associated with lung cancer risk in never smokers.
Yang	2010	Original		No convincing evidence to suggest an association between rs16969968 and lung cancer.
Pillai	2009	Original	COPD; Lung function	rs1051730 associated with COPD and lung function
Chen, Chen	2009	Original	ND; Symptoms of alcohol abuse	rs1051730 and rs16969968 associated with FTND score. Both SNPs were also associated with symptoms of alcohol abuse/dependence, but the associated alleles were the opposite of that of FTND. No association noted with cannabis abuse/dependence.
Schwartz	2009	Original	Lung cancer	rs1051730 associated with lung cancer risk in ever smoking African Americans (AAs) and whites. Associated risk increased following adjustment for CPD in ever smoking AAs, whereas risk estimate decreased (NS) after same adjustment in whites. No association noted with cancer in never smoking AAs or whites. CPD did *not* vary by rs105 genotype in AA sample.
Wang	2009	Original	Alcohol dependence	rs1051730 was associated with alcohol dependence, whilst rs16969968 was not.
Shiraishi	2009	Original	Lung cancer	Asian sample. rs16969968 and rs1051730 associated with lung cancer. Relationship observed for all histological types, and in both smokers and 'non-smokers' (latter group defined as never regular smokers).
Grucza	2008	Original	Cocaine dependence	The minor (A) allele of rs16969968 *protective* for cocaine dependence (whilst also a risk factor for ND). Effect replicated in additional sample.
Young	2008	Original	COPD	rs16969968 associated with COPD. Authors conclude that the association previously noted between this SNP and lung cancer could largely be explained through its relationship to COPD.
Liu	2008	Original	Lung cancer	rs1051730 associated with lung cancer.
Spitz	2008	Original	CPD; FTND; age at onset of lung cancer	No evidence of association between rs1051730 and lung cancer in never smokers. Partial sample overlap with Amos et al. Age at onset of lung cancer was modified by genotype (risk genotype = earlier age of onset). Highest risk for variant noted in *lightest* smokers. No

				association between rs1051730 and bladder/renal cancer.
Amos	2008	Original	Lung cancer	rs1051730 associated with lung cancer. In Texas population, adjusting for pack-years did not alter relationship. In UK population, this adjustment slightly weakened relationship.
Hung	2008	Original	Lung cancer	rs1051730 associated with lung cancer risk. rs16969968 also associated in 5 subsequent replication studies. Similar risk observed across all histological types. NO association observed between rs16969968 and head and neck cancers (including those of the oral cavity, larynx, oesophagus).
Thorgeirsson	2008	Original	Smoking quantity; ND; Lung cancer; Peripheral arterial disease	rs1051730 associated with lung cancer and peripheral arterial disease.

CPD = cigarettes per day, ND = nicotine dependence; FTND = Fagerström Test for Nicotine Dependence; COPD = chronic obstructive pulmonary disease; SNP = single nucleotide polymorphism; AA = African American; UADT = upper aerodigestive tract.

Table 1.3. Other phenotypes associated with rs1051730/rs16969968

Author	Year	Original research or meta-analysis	Phenotype/s associated with SNP/s	Notes
Janes	2011	Original	Smoking cue reactivity	FMRI study. rs16969968 associated with smoking-related cue reactivity in areas related to memory and habitual behaviour (dorsal striatum and hippocampus). NB: a*bsence* of the A allele associated with increased reactivity.
Winterer	2010	Original	ND; Cognition	rs1051730 and rs16969968 associated with ND and cognitive performance (cognitive domains from the WAIS-R, and n-back task performance).
Etter	2009	Original	Novelty seeking (marginal)	Potential association between rs16969968 and the temperament trait novelty seeking, although finding not robust to correction for multiple testing. No association noted with smoking status or cotinine levels.

FMRI = functional magnetic resonance imaging; ND = nicotine dependence; WAIS-R = Wechsler Adult Intelligence Scale (Revised).

1.4.2 Discovery of the association between rs16969968/rs1051730 and smoking behaviours

An association between rs16969968 in *CHRNA5* and nicotine dependence (ND) was first reported in 2007 in a candidate gene study conducted by Saccone and colleagues (Saccone et al., 2007), with the minor A allele found to confer increased risk. The following year, the same locus (tagged by rs1051730 in *CHRNA3*, a variant highly correlated with rs16969968) was also found to be associated with smoking quantity, this time identified in a GWA study conducted by Thorgeirsson and colleagues (Thorgeirsson et al., 2008). This study also demonstrated an association between rs1051730 and nicotine dependence and two smoking-related diseases, namely lung cancer and peripheral arterial disease. Notably, whilst the candidate gene study was published first, it was the GWA study that made much more of an impact. This may have been because *CHRNA5* was not recognised as a particularly strong candidate at the time, given the then known neurobiology of tobacco dependence - more emphasis had been placed on genes encoding α4 and β2 subunits which had been implicated by animal models as critical to the experience of nicotine's reinforcing effects (e.g., Picciotto et al., 1998). In contrast, the GWA study did not require a strong prior hypothesis regarding gene selection, as this approach is inherently agnostic with respect to candidacy. Furthermore, the simultaneous demonstration of an association between this locus and two smoking-related diseases lent further authority to this finding. These initial studies were followed by a number of others documenting a range of associations between this locus and smoking-related behaviours and diseases.

1.4.3 Phenotypes associated with rs16969968/rs1051730

Smoking behaviour. The 15q locus has primarily been associated with measures of heaviness of smoking, including ND and smoking quantity, although there is some evidence for other phenotypes.

SNPs rs1051730 and rs16969968 have been repeatedly associated with ND, typically assessed using the Fagerström Test for Nicotine Dependence (FTND) (L. S. Chen et al., 2009; X. Chen et al., 2009; Grucza et al., 2010; Johnson et al., 2010; Saccone, Saccone et al., 2009; Saccone, Wang et al., 2009; Saccone et al., 2007; Thorgeirsson et al., 2008; Wassenaar et al., 2011; Winterer et al., 2010). The impact of this locus on ND (and other smoking-related phenotypes) may be modified by different factors. The relationship has, for instance, been shown to be modified by age of smoking onset, although with inconsistent findings. Grucza et al. (2010) found that SNP rs16969968 exhibited a larger effect in late-onset smokers (post 16 years), whilst in contrast Weiss et al. (2008) noted an association between this locus and severity of nicotine dependence only in individuals who became regular smokers *before* the age of 16. Reasons underlying this disparity are unclear. A parsimonious explanation would be that these were chance findings. However, they do illustrate the potential importance of age of smoking onset, which is plausibly supported by research highlighting differential effects of nicotine exposure in adolescent and adult rats (e.g., Schochet, Kelley, & Landry, 2004).

Another related issue to be considered concerns the impact of these SNPs at different ages. Both Rodriguez et al. (2011) and Ducci et al. (2011) have sought to address this question, comparing the effects of this locus on smoking behaviour during adolescence and adulthood. Although phenotype definition and ages studied vary between these studies and are not directly comparable, both draw a similar

conclusion - the effect of this locus on smoking behaviour appears to be consistent during both adolescence and adulthood. Rodriguez et al. (2011) found that rs16969968 was associated with continued smoking in individuals who have experimented with tobacco, with similar effects noted at ages 13-15 years and at 18 years. Ducci et al. (2011) found that rs1051730 was associated with regular/heavy smoking, again with similar effects noted at ages 14 and 31 years.

Environmental factors have also been shown to impact upon the relationship between rs1051730/rs16969968 and smoking-related behaviours, such as parental monitoring (L. S. Chen et al., 2009), peer smoking (Johnson et al., 2010), and childhood adversity (Xie et al., 2011). This is discussed in more detail in Chapter 3.

Smoking quantity, typically assessed in terms of self-reported daily cigarette consumption, is also well established as a correlate of rs1051730 and rs16969968 genotypes (Breetvelt et al., 2011; Caporaso et al., 2009; Freathy et al., 2009; Kaur-Knudsen, Bojesen, Tybjaerg-Hansen, & Nordestgaard, 2011; Keskitalo et al., 2009; Lips et al., 2010; Marques-Vidal et al., 2011; Sarginson et al., 2011; Siedlinski et al., 2011; Sorice et al., 2011; Thorgeirsson et al., 2008; Wassenaar et al., 2011). Further, several meta-analyses published during the course of my Ph.D. have consistently documented this relationship (Furberg et al., 2010; Liu et al., 2010; Thorgeirsson et al., 2010). Each copy of the minor (risk) allele appears to account for approximately one cigarette per day in terms of variance in smoking quantity (Furberg et al., 2010). Given the above, it is perhaps unsurprising that levels of cotinine (the primary metabolite of nicotine) have also been found to associate with rs1051730 and rs16969968 genotype (Keskitalo et al., 2009; Le Marchand et al., 2008; Timofeeva et al., 2011). What is interesting, however, is that the relationship between this locus and nicotine metabolite levels appears to be stronger than the relationship noted

between this locus and daily cigarette consumption. Keskitalo et al.(2009) for instance found that rs1051730 was associated with both daily cigarette consumption and circulating cotinine levels, but, critically, also noted that the proportion of variance accounted for by this SNP was nearly five times greater for cotinine relative to daily cigarette consumption. This is explored in detail in Chapter 4.

Evidence for an association between rs1051730/rs16969968 and smoking cessation has been observed, although evidence for this relationship is weaker than that observed for ND and smoking quantity. Freathy et al. (2009) found an association between rs1051730 and reduced ability of women to quit smoking during pregnancy, an effect subsequently replicated by Thorgeirsson & Stefansson (2010). In further support, Munafò et al. (2011) found weak evidence of an association between rs1051730 and short-term cessation outcome in a combined analysis of two prospective clinical trial samples, although no evidence of association was noted at later follow-up. However, Breetvelt et al. (2011) and Lips et al. (2010) found no association between rs16969968 and smoking cessation, while Breitling et al. (2009) also failed to note an association between rs16969968 and rs1051730 and cessation, as assessed in ever heavy smokers (>20 cigs/day). In a similar vein, De Ruyck et al. (2010) found no association between rs1051730 and the presence of withdrawal symptoms or smoking cessation outcome following short-term nicotine patch treatment. Furthermore, Marques-Vidal et al. (2011) found no evidence for association between rs1051730 and willingness, attempt, or preparation to quit.

It is unclear whether or not rs1051730/rs16969968 is associated with smoking initiation. Lips et al. (2010) and Kaur-Knudsen et al. (2011) found no association between this locus and smoking initiation. A recent twin study (Maes et al., 2011) suggested that this locus plays a much more prominent role in ND relative

to smoking initiation/experimentation. However, Sherva et al. (2008), found an association between rs16969968 and smoking status (regular smoker vs. never smoker). Of particular interest, they also found an association between rs16969968 and positive first smoking experiences, specifically experience of a 'pleasurable buzz'. This may mediate the association between this SNP and increased risk of regular smoking. Inconsistencies in the definition of the 'initiation' phenotype may have hampered progress in this area – for example, the genes influencing initial experimentation (i.e., first puff) may differ from those underlying progression from experimentation to regular use. This is discussed at length in Chapter 3.

Cancer. Many diseases have been associated with SNPs rs16969968 and rs1051730, amongst which lung cancer is certainly the most frequently reported, and has been noted across a range of histology types (adenocarcinoma; squamous cell; large cell; small cell), and in European, Asian, and African American samples (Amos et al., 2010; Amos et al., 2008; Hung et al., 2008; Jaworowska et al., 2011; Kaur-Knudsen et al., 2011; Lips et al., 2010; Liu et al., 2008; Saccone et al., 2010a; Sakoda et al., 2011; Schwartz, Cote, Wenzlaff, Land, & Amos, 2009; Shiraishi et al., 2009; Spitz, Amos, Dong, Lin, & Wu, 2008; Timofeeva et al., 2011; Truong et al., 2010; Wang et al., 2010; Wassenaar et al., 2011; although see Yang et al., 2010). There is considerable debate as to whether this association is direct or mediated via the variants' association with smoking quantity. Briefly, the former (direct) argument is supported by studies demonstrating a relationship between this locus and cancer *following adjustment for smoking quantity* (e.g., Kaur-Knudsen et al., 2011; Wassenaar et al., 2011), whilst the latter (indirect) is supported by studies which fail to note an association between this locus and cancer in never smokers (e.g., Girard et al., 2010), and the inadequacy of self-reported smoking measures in capturing true

tobacco exposure (Munafo et al., 2012). This is discussed in detail in Chapter 4. Several lung cancer specific phenotypes have also been associated with this locus, age of cancer onset/diagnosis being most predominantly reported (Lips et al., 2010; Sakoda et al., 2011; Spitz et al., 2008; Truong et al., 2010) – presence of the minor allele is consistently associated with earlier age of onset/diagnosis (although see Jaworowska et al., 2011). SNP rs1051730 has also been associated with larger tumour size at diagnosis for squamous cell carcinoma (X. Chen et al., 2011). However, it does not appear to be associated with survival time in lung cancer patients (Xun et al., 2011). Additional cancers linked to this locus include upper aerodigestive tract cancers (e.g., those of the oral cavity, larynx, oesophagus) (Lips et al., 2010), although this association has not been consistently shown (Hung et al., 2008), and more recent work suggests that it may be limited to women only (D. Chen et al., 2011). Bladder cancer has also been associated with this locus (Gago-Dominguez et al., 2011; Kaur-Knudsen et al., 2011), although, again, this finding has not been consistently shown (Jaworowska et al., 2011; Spitz et al., 2008). Finally, Chen et al. (2011) found no association between rs1051730 and pancreatic cancer risk.

Alcohol and substance use. Alongside tobacco dependence, rs1051730 and rs16969968 have been linked to dependence upon other drugs of abuse, including opiates (Erlich et al., 2010), cocaine (Grucza et al., 2008), and alcohol (X. Chen et al., 2009; Wang et al., 2009). Erlich et al. (2010) found that the minor allele of rs16969968 was associated with opioid dependence severity, the same allele that has consistently been associated with ND. In contrast, Grucza et al. (2008) found this same minor allele to be *protective* for cocaine dependence. Similarly, Chen et al. (2009) found that the major alleles of rs16969968 and rs1051730 were associated

with symptoms of alcohol abuse/dependence, whilst simultaneously demonstrating an association between the minor alleles and ND. They found no evidence for an association between these variants and cannabis dependence. Whilst these opposing effects are intriguing, they are based on a very limited number of studies and therefore require replication.

Other disease outcomes. Chronic obstructive pulmonary disease/emphysema, a common smoking-related disease, has also been associated with rs16969968 and rs1051730 (Kaur-Knudsen et al., 2011; Kim et al., 2011; Lambrechts et al., 2010; Pillai et al., 2009; Wang et al., 2010; Young et al., 2008). Arguments as to whether this association is direct or mediated via the association with smoking quantity are also common here. An association between this locus and cardiovascular disease has also been demonstrated. For instance, Thorgeirsson et al. (2008) observed an association between rs1051730 and peripheral arterial disease, also a known smoking-related disease. Finally, Hong et al. (2011) have demonstrated an association between rs16969968 and schizophrenia.

Other non-disease outcomes. How do we explain the associations noted between SNP rs16969968/rs1051730 and smoking related behaviours? Several studies investigating associations between these variants and cognitive and personality related phenotypes offer some insight. Etter et al. (2009) found marginal evidence of an association between rs16969968 and novelty seeking. Individuals with the AA (ND risk) genotype had higher novelty seeking scores than individuals of GG or AG genotype, suggesting mediation by personality trait (of note, however, no association was observed with ND). Winterer et al. (2010) reported an association between both rs1051730 and rs16969968 and cognitive performance as assessed by the Wechsler-Adult-Intelligence Scale and an n-back task measure of executive

function. The alleles associated with lower cognitive performance were also those associated with increased risk for ND. Against a background of previous research highlighting the role of nicotine as a cognitive enhancer (Warburton, 1992), the authors postulate that this locus may indirectly increase a subject's liability to ND as a result of cognitive augmentation by nicotine consumption. Indeed, the increased prevalence of smoking noted in samples of individuals with neurocognitive disorders (e.g., attention-deficit hyperactivity disorder) has been attributed to nicotine's beneficial effect on cognitive performance (e.g., improving attention) (Sacco, Bannon, & George, 2004). It has also been proposed that genetic effects on smoking behaviours may be mediated in part by their effect on reactivity to smoking cues. Janes et al. (2011) found an association between rs16969968 and brain reactivity to smoking-related cues assessed by functional magnetic resonance imaging (fMRI). They found that women *without* the risk allele for ND showed greater reactivity to smoking cues in regions such as the hippocampus and dorsal striatum relative to women possessing this allele. The authors speculate that smokers *without* the ND risk allele may thus continue to smoke due to heightened cue reactivity. The results of this study are counter-intuitive in comparison to previous research. However, differences in nicotine dependence were controlled for when comparing smokers with and without the ND risk allele. Other studies have not done this when investigating the effects of this variant, which may partly explain these results. However, the sample size was small, which increases the possibility that statistically significant results may reflect false positives (Green et al., 2008), and so these results should be interpreted with particular caution until they have been replicated.

1.4.4 Determining linking mechanisms

The evidence linking variants rs1051730 and rs16969968 to smoking-related behaviours is, at the time of writing, compelling. What is less clear, however, is the fundamental mechanism linking the two. Exactly how do these polymorphisms exert their effect? Let us first consider their functional significance. SNP rs1051730 in *CHRNA3* is a coding**,** synonymous variant (http://genome.ucsc.edu/), i.e., a variant which does not result in an amino acid change in the subsequent protein, which is therefore unlikely to be of functional significance. This variant may act as a proxy or tag for a functional SNP however, which may underlie the observed associations. In contrast to rs1051730, rs16969968 in *CHRNA5* is a missense mutation, resulting in an amino acid change (aspartate to asparagine) in the resultant α_5 nAChR subunit protein. This variant is of definite functional significance – in vitro studies have demonstrated that α_5 receptor complexes with the aspartic acid variant exhibit a twofold greater maximal response to a nicotine agonist compared to α_5 receptor complexes containing the asparagine variant (i.e., the risk variant robustly associated with ND) (Bierut et al., 2008). Building upon this foundation of research, Fowler et al. (2011) sought to establish the underlying mechanism through an elegant series of experiments involving α_5 knockout mouse models (analogous to individuals with reduced α_5 receptor function, i.e., carriers of the rs16969968 risk allele). They noted that knockout mice responded more vigorously than wild-type mice for nicotine infusions at high doses. Whilst wild-type mice appeared to titrate delivery of nicotine dose (through self-administration) to achieve a consistent, desired level, knockout mice did not, consuming greater amounts as dosage increased. This led the authors to propose that deficient α_5 signalling attenuates the negative effects of nicotine that serve to limit its intake, a conclusion which fits well with human research (i.e.,

smokers carrying the rs16969968 risk allele are likely to smoke more heavily than their counterparts without the risk allele). Furthermore, they also demonstrated that this effect could be 'rescued' in α_5 knockout mice through injection of a lentivirus vector into the medial habenula (MHb), rescuing expression of α_5 subunits in this region. The knockout mice did not appear to differ from wild-type mice in experience of the rewarding effects of nicotine, but the inhibitory effect of high nicotine doses on the activity of reward circuitries observed in wild-types appeared to have been largely abolished in knockout mice. This observation is complemented by a previous study by Jackson (2010), where the differential effects of nicotine dose on reward between α_5 knockouts and wild-types was illustrated using a conditioned place preference task. Fowler et al. (2011) further determined that this effect appeared to be mediated via the pathway between the MHb and the interpeduncular nucleus (IPN, to which the MHb projects) through α_5 containing nAChRs. Diminished IPN activity in response to nicotine was observed in knockouts, and additionally, disruption of IPN activity increased nicotine self-administration. In short, it appears that high doses of nicotine stimulate the MHb-IPN tract through nAChRs containing α_5 subunits. This results in the relay of an inhibitory motivational signal serving to limit further drug intake. This pathway acts alongside the classic 'reward' pathway.

1.5 Summary

Tobacco remains the leading preventable cause of death worldwide. In order to address this epidemic, it is important that we have a thorough understanding of the aetiology of tobacco use and dependence. Tobacco dependence is a complex disease. Both genetic and environmental factors contribute to its aetiology. This has been demonstrated by both twin and adoption studies. The advent of genome-wide

technologies has greatly facilitated the search to determine which specific genetic factors contribute to tobacco use phenotypes. A locus within the nAChR gene cluster *CHRNA5-A3-B4* has generated particular interest – that marked by SNPs rs16969968 in *CHRNA5* and rs1051730 in *CHRNA3*. Research is now focused on determining exactly how these variants influence tobacco use phenotypes, and on identifying additional variants robustly associated with these phenotypes.

1.6 Aims

1.6.1 Overview

The primary aim of this thesis was to determine the role played by the *CHRNA5-A3-B4* gene cluster in smoking-related behaviours, with an emphasis on phenotype refinement to aid understanding of the mechanisms underlying these associations. Multiple approaches have been utilised to address this objective, namely systematic review and meta-analysis, genetic epidemiology, and laboratory-based techniques. A secondary objective was to identify additional genetic variants robustly associated with smoking-related phenotypes. The clinical utility of this line of research is to further our understanding of the genetic contribution of smoking-related behaviours, which may ultimately enable us to improve and personalise smoking cessation treatments. This, in turn, may help to reduce the substantial health concern associated with tobacco use.

1.6.2 Systematic review and meta-analysis (Chapter 2)

Through systematic review and meta-analysis, I sought to evaluate the strength of evidence for the association between rs1051730 (*CHRNA3*) and rs16969968 (*CHRNA5*) and heaviness of smoking, assessed in terms of daily cigarette consumption. Secondary aims were to determine which (if either) of the two

variants provided a stronger genetic signal, test for the existence of small study bias, explore the impact of year of publication, and investigate the impact of ancestry and disease state as potential moderating variables.

1.6.3 Genetic epidemiology (Chapter 3)

Using data from a prospectively assessed cohort, I sought to determine the association between rs1051730 and smoking initiation, given somewhat conflicting evidence in the field regarding this relationship. Initiation was assessed both in terms of 'ever use' of cigarettes, and also in terms of initial smoking trajectories – a novel, sophisticated phenotype which captures the complexities of initial cigarette use, determined using repeated measures of smoking frequency. Given that this variant is located in a gene responsible for encoding the α_3 nicotinic receptor subunit, I hypothesised that rs1051730 would not be associated with ever (i.e., very first) use of cigarettes, but may be associated with initial smoking trajectories, namely those capturing progression from initial exposure to tobacco through to regular use. I also sought to determine potential modification in the expression of these genetic effects by a well-established environmental risk factor for smoking initiation, namely parental monitoring. Given that this variable may plausibly influence the probability of exposure to tobacco/opportunities to smoke, I hypothesised that level of parental monitoring would moderate the expression of the predicted genetic effect.

1.6.4 Laboratory-based techniques (Chapter 4)

Previous research denotes a much stronger association between rs1051730 and cotinine level (a precise, objective assessment of tobacco exposure) relative to self-reported daily cigarette consumption. Moreover, this relationship appears robust to adjustments made for daily cigarette consumption (Munafo et al., 2012). This suggests that even among *equal* cigarette consumers, there is genetically influenced

variation in total tobacco exposure. Building upon this foundation of research, I sought to determine whether differences in smoking topography (i.e., how a cigarette is smoked) influence the relationship between rs1051730 and salivary cotinine level.A recall-by-genotype based approach was proposed to address this objective.

1.6.5 Genome-wide meta-analysis (Chapter 5)

The primary objective of this project was to identify additional genetic variants robustly associated with heaviness of smoking, in this instance assessed in terms of cotinine level, which, as has previously been demonstrated, is a more precise and objective assessment of tobacco exposure relative to self-reported daily cigarette consumption, thus offering a 'cleaner' genetic signal, and maximising power to detect genetic effects. To this end, we created a consortium ('Cotinine Consortium') comprised of seven studies to conduct a genome-wide meta-analysis of this phenotype. Given that this approach is agnostic with respect to candidacy, no *a priori* hypotheses were required.

Chapter 2

Systematic review and meta-analysis:

Association of the *CHRNA5-A3-B4* gene cluster with heaviness of smoking

2.1 Introduction

2.1.1 Meta-analysis

Meta-analysis is a statistical technique which allows the results from two or more separate studies to be combined (Deeks, Higgins, & Altman, 2011). Particular strengths of meta-analyses include "an increase in power, an improvement in precision, the ability to answer questions not posed by individual studies, and the opportunity to settle controversies arising from conflicting claims" (Deeks et al., 2011). In contrast to the classic narrative review, which is subjective and prone to bias and error, meta-analysis is based on rigorous systematic review methodology, allowing a more objective appraisal of research evidence (Egger & Smith, 1997).

Central to the systematic review (including meta-analysis) is clear and transparent methodology for both identifying relevant studies for inclusion, and extracting and analysing information from them (Kirkwood & Sterne, 2003). With regards to study inclusion for meta-analytic purposes, "completeness and combinability" of evidence is key (Egger, Smith, & Phillips, 1997).

One problem inherent to meta-analyses based solely on the published literature is that of publication bias (i.e., preferential publication of significant findings over null findings). Publication bias, more accurately described as "small study bias" - a term used to "describe a trend for the smaller studies in a meta-analysis to show larger treatment effects" (Sterne, Gavaghan, & Egger, 2000) - can distort combined effect estimates. Effect size estimates from small studies tend to

vary widely. For an effect borne out of a small study to be significant, it will have to be very large. Non-significant small studies are unlikely to be published. Those that are significant are likely to feature very large effect sizes. As a consequence of the above, inclusion of such studies in meta-analyses is likely to skew the overall effect size estimate, leading to an over-estimate of said effect (Sterne et al., 2000). Fortunately, however, the presence of such bias can be formally tested and corrected for using statistical tests (Egger, Davey Smith, Schneider, & Minder, 1997). An additional cause of bias in meta-analysis is that resulting from poor study quality (Kirkwood & Sterne, 2003). Given the inherent difficulties in objective assessment of study quality, Egger, Smith et al. (1997) suggest the use of sensitivity analysis to examine the potential impact of study quality on results, over and above the finite exclusion of certain studies on the basis of a somewhat subjective assessment.

In calculating overall effect estimates, meta-analytical methods adopt a weighted approach, whereby the results from larger studies (with smaller standard errors) are given greater weighting than those from smaller studies (with larger standard errors). A "fixed effects" or "random effects" model is typically employed for calculating summary estimates in meta-analyses. Selection of the most appropriate model is dependent on the degree of heterogeneity observed between studies. This can be calculated using the I^2 statistic (see Higgins, Thompson, Deeks, & Altman, 2003)

In addition to providing single, combined effect estimates summarising defined literatures, meta-analyses are also able to answer questions beyond those posed by their constituent studies. For instance, they allow examination of effect size variation between subgroups (e.g., groups defined by gender, age, ethnicity, and so on) (Egger & Smith, 1997).

2.1.2 The CHRNA5-A3-B4 gene cluster and heaviness of smoking

As previously discussed, many original studies have documented an association between SNPs rs1051730 and rs16969968 in the *CHRNA5-A3-B4* gene cluster and heaviness of smoking. An association between the rs16969968 variant and smoking quantity was also recently identified in a meta-analysis of primarily new, unpublished data (Saccone et al., 2010b). Moreover, three recent genome-wide meta-analyses have further highlighted an association between this locus and smoking quantity (Furberg et al., 2010; Liu et al., 2010; Thorgeirsson et al., 2010).

The *CHRNA5-A3-B4* cluster is clearly an established region for smoking behaviour. At the time of project conception, no study-level meta-analysis had been conducted to determine the strength of association between rs1051730 specifically and smoking quantity, to determine whether the strength of association between both variants and smoking quantity differed according to sample ancestry or disease status, or to assess small study bias. It is of note that typically only one of these two SNPs tends to be used in analyses in the wider literature for practical reasons. It is therefore of interest to determine whether they should continue to be used interchangeably. We sought to: 1) evaluate the strength of evidence for the association between the rs16969968 and rs1051730 SNPs and heaviness of smoking (both in pooled and independent analyses), as measured by daily cigarette consumption, using meta-analytic techniques to synthesise existing published data; 2) explore which variant provides a stronger genetic signal; 3) test for the possibility that small study sample sizes may have biased findings; 4) explore the impact of year of publication; 5) investigate the impact of ancestry and disease state as potential moderating variables.

2.2 Methods

2.2.1 Selection of studies for study inclusion

Studies were included that reported data on the *CHRNA5* polymorphism rs16969968 and/or the *CHRNA3* polymorphism rs1051730 and smoking quantity. If data regarding smoking quantity were presented categorically, or were available but not reported by genotype, we contacted the authors to determine whether data in an appropriate format for inclusion were available. Three attempts were made to contact study authors. If these attempts did not result in the provision of data, the study was included but coded as 'data not available'. Studies in any language, reporting data on samples of any ethnic origin were included, as were studies reporting data on either single-sex samples or samples including both males and females.

Studies were excluded if no data on smoking quantity were available, neither of the SNPs of interest was investigated, or if extreme smoking quantity phenotypes had been selected for analysis. Reviews, letters to the editor and editorials were excluded if these did not present new or relevant data. Family based studies were also excluded, due to differences in analytical approach. Additionally, studies were excluded if an inappropriate study design was employed (e.g., DNA pooling).

2.2.2 Search strategy

The search was performed in Scopus and PubMed. These databases were searched from the first date available in each database up to 12th May 2010 using the following search terms: "*CHRNA5* or *CHRNA3* or *CHRNB4*"; "rs16969968 or rs1051730"; "smok* and 15q2*". Once articles had been collected, references were hand searched for additional studies of interest.

The titles and abstracts of studies identified by these search strategies were examined, and those clearly fitting the inclusion or exclusion criteria were retained or

excluded respectively. Of the remaining studies a more thorough examination of the full-text and supplementary material (if available) was required to determine retention or rejection. All duplications were deleted. Where studies reported previously published data we included data from only one of the publications, namely that reporting the largest sample. Ten per cent of all studies identified by the search strategy were additionally assessed for eligibility by a second reviewer (inter-rater agreement > 90%). Disagreements between reviewers were resolved by mutual consent.

2.2.3 Data extraction

For each study the following data were extracted: a) authors and year of publication, b) sample characteristics (ancestry; disease state), c) SNP(s) studied, and d) mean (*M*), standard deviation (*SD*) and sample size (*N*) for cigarettes per day by genotype. Genotype frequencies were used to calculate deviation from Hardy-Weinberg equilibrium (HWE), i.e., deviation of population allele and genotype frequencies from a constant state of equilibrium. Deviations from HWE may be due to a number of factors such as population admixture/stratification, selection at a locus or inbreeding, but are most frequently due to genotyping error, which can greatly reduce the power of a genetic study (Leal, 2005; Sen & Burmeister, 2008). Ancestry was coded as European or 'Other', given the paucity of studies reporting data on non-European samples. To be coded as European, a sample had to be comprised of at least 95% European individuals.

2.2.4 Data analyses

Given the high linkage disequilibrium between rs16969968 and rs1051730 (European: $r^2 = 0.902$, Japanese/Chinese: $r^2 = 1.000$, African: r^2 = unavailable; calculated using HapMap data in conjunction with SNAP

[http://www.broadinstitute.org/mpg/snap/ldsearchpw.php]), we initially conducted pooled analyses incorporating data from all samples, regardless of SNP studied, omitting one dataset if data on both SNPs had been collected for a sample. The standard additive model of genetic action was used for evaluation. Small study bias was assessed using the Egger test (Egger, Davey Smith et al., 1997) for both pooled and independent SNP analyses. The impact of year of publication on effect size estimate was also examined. Data were analysed within both a fixed- and random-effects framework. Individual study effect sizes were pooled to generate a summary effect estimate and 95% *CI*, the significance of which was determined using a Z test. Stratified analyses by sample ancestry (European vs. Other) and disease state (control/population vs. disease/partial disease, e.g., lung cancer cases) were conducted to ascertain the potential moderating effects of these variables. We also explored which SNP provided a stronger genetic signal. The differences in pooled effect sizes were determined using a Z test.

Between-study heterogeneity was examined using a chi-square test, and quantified through calculation of I^2 - the conventional bounds for low, medium and high heterogeneity based on the I^2 statistic being 25%, 50% and 75%, respectively. Data were analysed with Comprehensive Meta-Analysis Version 2 statistical software (Biostat, Englewood NJ).

2.3 Results

2.3.1 Study selection

The search of Scopus and PubMed databases provided 585 records. Two additional records were identified through other sources (hand-searching references of identified papers). After adjusting for duplications, 432 records remained. Of these, 325 were discarded because after reviewing the abstracts it appeared that these

papers clearly did not meet the required criteria. The full texts of the remaining 107 studies were examined in detail (Figure 2.1). Of these, 37 were identified for inclusion in the meta-analysis (Tables 2.1 and 2.2).

Figure 2.1. Flow diagram of study selection.

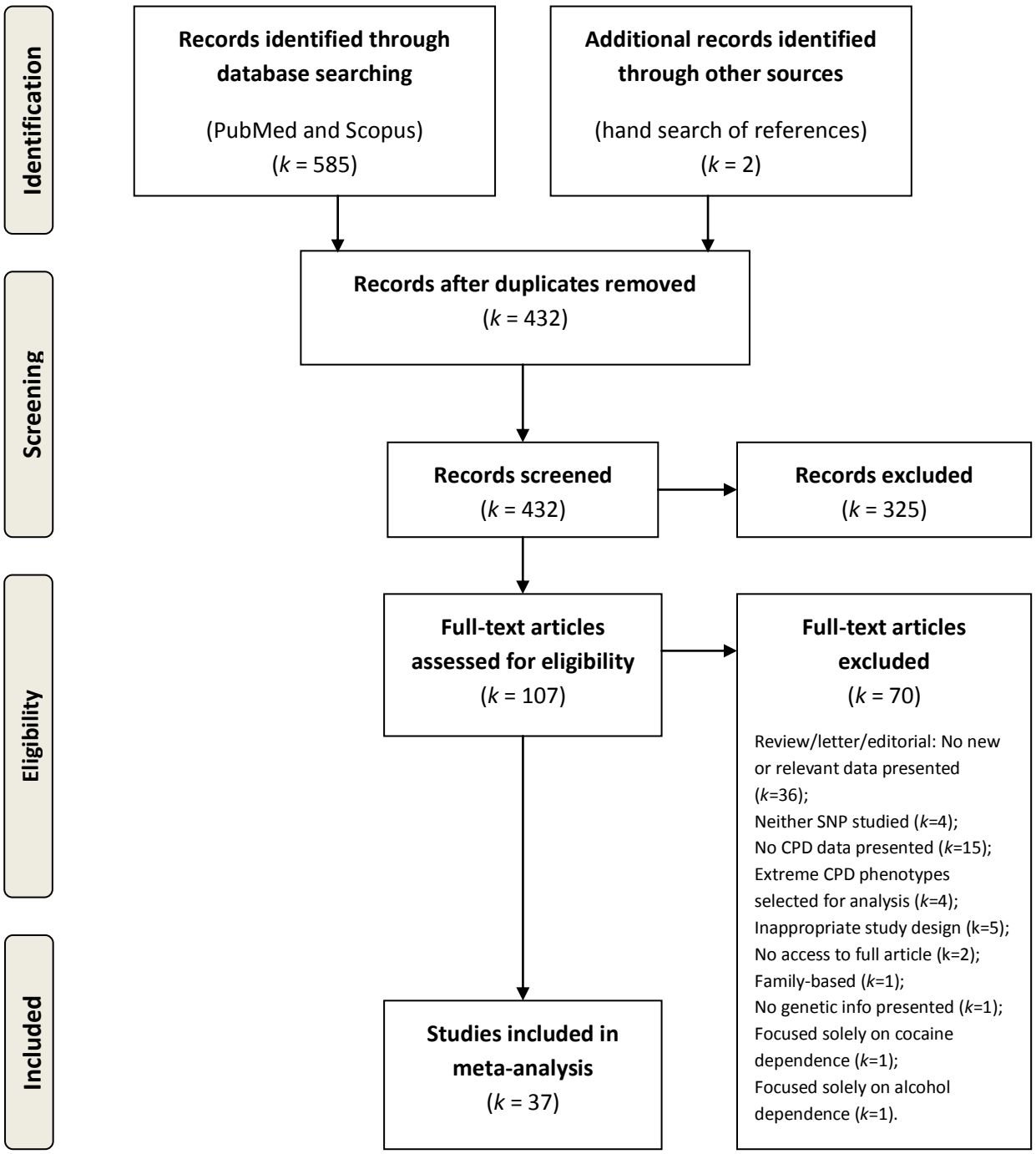


Table 2.1. Studies included in rs16969968 meta-analysis

Study	Year	Cigarettes per Day									HWE	Ancestry	Disease State	Duplicate
		GG			GA			AA						
		M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*				
Baker (WI/UTAH)	2009	Data not available. Data also analysed elsewhere (Weiss et al., 2008)												
Breitling	2009	25.51	10.70	200	25.28	10.63	241	28.37	10.83	89	Yes	European	No	Yes
Broderick (Phase II cases)[1]	2009	20.80	13.20	861	21.60	13.30	1033	23.00	13.60	332	Yes	European	Yes	Yes
Broderick (Phase II controls)[1]	2009	18.20	12.60	423	18.50	10.40	391	18.60	10.70	93	Yes	European	No	Yes
Caporaso	2009	Data not available												
Chen (VAANX-ND)	2009	26.24	16.98	374	27.09	16.72	351	29.32	15.68	73	Yes	European	Yes	Yes
Chen (VAFTND)	2009	24.99	15.00	461	27.28	16.25	499	29.51	17.21	142	Yes	European	Yes	Yes
Etter	2009	26.70	19.20	62	23.60	10.90	117	21.90	10.50	32	Yes	European	No	No
Greenbaum[2]	2009	8.73	7.02	47	8.89	6.62	62	12.31	9.09	16	Yes	Other	No	No
Grucza (COGA, nonhabitual smokers)	2008	8.47	4.08	43	8.52	4.48	46	11.90	4.56	10	Yes	European	Yes	No
Grucza (COGA, habitual smokers)	2008	25.72	9.84	139	27.47	11.48	156	26.16	10.43	45	Yes	European	Yes	No
Grucza (FSCD, nonhabitual smokers)	2008	12.12	7.25	74	13.76	9.28	51	10.29	6.63	7	Yes	European	Yes	No
Grucza (FSCD, habitual smokers)	2008	21.64	8.00	61	20.11	6.58	54	24.47	9.26	19	Yes	European	Yes	No
Hung	2008	Data analysed elsewhere (Lips et al., 2009)												
Landi (CPSII)[3]	2009	9.21	12.45	60	9.86	14.48	71	11.56	16.30	17	Yes	European	Yes	Yes
Le Marchand (Hawaii)	2008	22.03	10.12	393	23.54	9.80	159	27.57	10.91	31	No	Other	No	Yes

Le Marchand (UMN-1)	2008	24.00	5.65	41	27.50	7.85	48	25.50	8.64	10	Yes	European	No	Yes
Le Marchand (UMN-2)[4]	2008	19.90	8.20	80	21.60	8.95	47	22.60	5.32	21	No	Other	No	Yes
Lips (Central Europe)	2009	14.10	7.91	1503	14.30	10.81	1795	14.50	8.19	526	Yes	European	Yes	No
Lips (Toronto)	2009	17.50	13.17	151	16.60	12.32	180	19.20	13.35	50	Yes	European	Yes	No
Lips (EPIC)	2009	12.80	8.88	841	13.00	8.14	1018	13.60	8.63	353	Yes	European	Yes	No
Lips (Liverpool)	2009	14.30	10.57	429	14.90	11.25	402	16.80	11.72	109	Yes	European	Yes	No
Lips (Hunt/Tromso)	2009	9.60	4.97	148	10.90	5.28	167	11.60	5.95	42	Yes	European	Yes	No
Lips (ARCAGE)	2009	15.60	14.21	958	15.30	14.10	1194	16.00	11.27	339	Yes	European	Yes	No
Lips (Latin America)	2009	14.50	17.01	1112	15.30	16.04	817	17.80	12.86	176	Yes	European	Yes	No
Liu (GELCC, cancer cases and controls)	2008	Data not available												
Liu	2010	Data not available												
McKay	2008	Data analysed elsewhere (Lips et al., 2009)												
Pillai (Bergen Discovery, cases and controls)	2009	Data not available												
Ray (Discovery)	2010	Data not available												
Sherva	2008	Data not available												
Shiraishi (cases)	2009	28.70	14.00	919	29.70	11.80	64	25.00	2.20	2	Yes	Other	Yes	Yes
Shiraishi (controls)	2009	20.00	12.20	349	27.50	10.20	11	15.00	NA	1	No	Other	No	Yes
TAG	2010	Data not available												
Thorgeirsson	2010	Data not available												
Weiss (Utah/WI/LHS)	2008	Data not available												

Wu	2009	Data not available											
Yang	2010	Data not available											
Young (Lung cancer cases)	2008	19.53	10.51	157	19.59	8.42	180	19.47	7.87	64	Yes	European	Yes	Yes
Young (COPD cases)	2008	Data not available											
Young (control smokers)	2008	22.23	9.53	222	24.14	11.17	202	27.87	15.36	45	Yes	European	No	Yes
Young	2009	Data analysed elsewhere (Young et al., 2008)											
Zienolddiny (cases)	2009	14.84	9.63	112	14.26	7.94	186	15.77	9.20	59	Yes	European	Yes	Yes
Zienolddiny (controls)	2009	14.16	6.23	174	14.71	6.51	194	14.97	5.41	58	Yes	European	No	Yes

Disease state: no=population/control sample; yes=disease/partial disease sample. Duplicate: no=sample provided data for rs16969968 only and was included in pooled SNP analyses; yes=sample also provided data for rs1051730 and was excluded in pooled data analyses.

[1] rs16969968 not genotyped in Phase I cases and controls.

[2] These data are identical to rs1051730 data from Greenbaum et al. (2006). rs16969968 was not genotyped in Greenbaum et al. (2006) but was in Greenbaum et al. (2009). Authors suggested use of rs1051730 data for rs16969968 analysis given their complete correlation ($R^2=1$) in their primary sample.

[3] rs16969968 data not available for EAGLE, ATBC and PLCO samples.

[4] Data from additional participants provided by authors.

Table 2.2. Studies included in rs1051730 meta-analysis.

Study	Year	Cigarettes per Day									HWE	Ancestry	Disease state	Duplicate
		CC			CT			TT						
		M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*				
Amos (Texas Discovery, controls, current)	2008	23.00	12.20	207	23.70	11.50	221	24.40	11.00	52	Yes	European	No	No
Amos (Texas Discovery, controls, former)	2008	28.00	14.70	294	28.70	16.10	290	33.30	15.60	73	Yes	European	No	No
Amos (Texas Discovery, cases, current)	2008	27.80	13.20	194	28.40	11.70	262	30.40	14.60	95	Yes	European	Yes	No
Amos (Texas Discovery, cases, former)	2008	27.10	13.90	230	26.90	14.50	279	31.00	14.90	93	Yes	European	Yes	No
Amos (Texas Replication, controls, current)	2008	21.40	13.30	146	20.90	10.20	136	25.00	11.10	42	Yes	European	No	No
Amos (Texas Replication, controls, former)	2008	25.90	15.70	120	29.30	17.70	124	35.20	14.70	26	Yes	European	No	No
Amos (Texas Replication, cases, current)	2008	27.50	15.30	116	26.10	11.70	142	28.30	13.10	51	Yes	European	Yes	No
Amos (Texas Replication, cases, former)	2008	26.20	14.50	143	26.30	12.60	188	28.30	11.90	62	Yes	European	Yes	No
Amos (UK Replication, controls, current)	2008	16.90	8.40	159	18.20	8.60	157	15.50	8.90	34	Yes	European	No	No
Amos (UK Replication, controls, former)	2008	18.60	14.10	286	18.50	11.50	261	20.10	12.10	59	Yes	European	No	No
Amos (UK Replication, cases, current)	2008	22.60	10.70	174	22.00	11.80	280	23.60	10.20	97	Yes	European	Yes	No
Amos (UK Replication, cases, former)	2008	21.50	13.10	513	22.90	13.00	568	24.90	15.10	198	Yes	European	Yes	No
Baker (WI/UTAH)	2009	Data not available. Data also analysed elsewhere (Weiss et al., 2008)												
Breitling	2009	25.57	10.77	197	25.26	10.67	239	28.71	10.40	89	Yes	European	No	No
Broderick (Phase I cases)$_1$	2009	19.00	14.80	709	21.30	14.10	913	23.20	14.90	330	Yes	European	Yes	No
Broderick (Phase II cases)	2009	20.80	13.10	859	21.60	13.40	1036	22.90	13.40	332	Yes	European	Yes	No

Broderick (Phase II controls)	2009	18.20	12.60	423	18.50	10.40	391	18.60	10.70	93	Yes	European	No	No
Caporaso	2009	Data not available												
Chen (VAANX-ND)	2009	26.07	16.86	380	27.19	16.54	359	28.63	15.44	67	Yes	European	Yes	No
Chen (VAFTND)	2009	24.98	14.97	464	27.08	16.25	506	29.21	17.26	139	Yes	European	Yes	No
Conti	2008	Data not available												
Freathy (ALSPAC - third trimester data)$_2$	2009	10.44	6.96	522	10.93	6.43	561	11.45	7.13	157	Yes	European	No	No
Greenbaum	2006	8.73	7.02	47	8.89	6.62	62	12.31	9.09	16	Yes	Other	No	No
Greenbaum	2009	Data analysed elsewhere (Greenbaum et al. 2006)												
Hung (IARC)	2008	Data analysed elsewhere (McKay et al., 2008; Lips et al., 2009)												
Keskitalo	2009	15.80	8.78	194	17.67	9.90	221	17.42	8.20	60	Yes	European	Yes	No
Lambrechts (LEUVEN)	2010	22.20	9.80	259	23.40	10.40	301	23.40	10.60	99	Yes	European	Yes	No
Lambrechts (COPACETIC)	2010	19.80	9.20	200	21.40	10.60	203	23.60	9.80	53	Yes	European	Yes	No
Landi (EAGLE)	2009	14.49	12.64	1295	15.43	12.51	1898	17.37	13.08	701	Yes	European	Yes	No
Landi (PLCO)	2009	17.83	14.84	1360	19.70	15.21	1501	20.94	16.01	415	Yes	European	Yes	No
Landi (ATBC)	2009	20.20	9.01	1210	21.17	8.69	1405	22.52	9.51	386	Yes	European	Yes	No
Landi (CPSII)	2009	15.69	14.85	542	16.89	15.77	626	18.96	16.25	202	Yes	European	Yes	No
Le Marchand (Hawaii)	2008	22.01	10.11	393	23.37	9.84	157	27.78	10.89	31	No	Other	No	No
Le Marchand (UMN-1)	2008	24.10	5.68	40	27.50	7.85	48	25.50	8.64	10	Yes	European	No	No
Le Marchand (UMN-2)$_3$	2008	19.90	9.04	79	21.60	8.52	48	22.60	8.17	21	No	Other	No	No
Liu (GELCC, cancer cases and controls)	2008	Data not available												

Liu	2010	Data not available												
McKay	2008	Data not available												
Pillai	2009	Data not available												
Ray (Discovery)	2010	Data not available												
Rigbi	2008	Data analysed previously (Greenbaum et al. 2006)												
Schwartz (EA controls)	2009	17.11	10.36	177	18.89	10.81	168	18.54	9.69	52	Yes	European	No	No
Schwartz (EA cases)[4]	2009	25.71	11.25	190	28.31	13.74	273	29.29	13.67	86	Yes	European	Yes	No
Schwartz (AA controls)	2009	14.90	11.13	194	14.28	9.19	39	5.67	4.16	3	Yes	Other	No	No
Schwartz (AA cases)[4]	2009	16.66	10.77	267	20.31	11.90	96	22.50	8.64	8	Yes	Other	Yes	No
Shiraishi (cases)	2009	28.70	14.00	922	29.70	11.80	61	25.00	2.20	2	Yes	Other	Yes	No
Shiraishi (controls)	2009	20.20	12.30	350	25.20	10.20	10	15.00	NA	1	No	Other	No	No
Spitz (Bladder cancer cases)[5]	2008	37.58	29.42	299	45.17	30.38	314	48.14	30.10	81	Yes	European	Yes	No
Spitz (Bladder controls)	2008	29.78	30.37	224	30.37	26.30	198	38.44	27.60	57	Yes	European	No	No
Spitz (Renal cancer cases)	2008	29.72	27.59	72	29.12	21.66	63	19.04	12.78	7	Yes	European	Yes	No
Spitz (Renal controls)	2008	34.56	31.88	82	30.14	31.10	61	41.35	35.51	15	Yes	European	No	No
TAG	2010	Data not available												
Thorgeirsson (Iceland/Spain/Netherlands)	2008	Data not available												
Thorgeirsson	2010	Data not available												
Weiss (Utah/WI/LHS)	2008	Data not available												
Wu	2009	Data not available												

Young (Lung cancer cases)	2008	19.53	10.51	157	19.59	8.42	180	19.47	7.87	64	Yes	European	Yes	No
Young (COPD cases)	2008	21.09	8.86	161	23.20	11.05	204	24.44	13.97	54	Yes	European	Yes	No
Young (control smokers)	2008	22.23	9.53	222	24.14	11.17	202	27.87	15.36	45	Yes	European	No	No
Zienolddiny (cases)	2009	14.66	9.55	110	14.19	8.03	184	16.15	9.08	58	Yes	European	Yes	No
Zienolddiny (controls)	2009	14.13	6.28	174	14.78	6.47	195	15.05	5.47	56	Yes	European	No	No

Disease state: no=population/control sample; yes=disease/partial disease sample.

[1] rs1051730 not genotyped for Phase I controls.

[2] All EFSOCH and ALSPAC lifetime, pre-pregnancy and first-trimester CPD data available in categorical format only. CPD data available in continuous format for ALSPAC third-trimester only (displayed).

[3] Data from additional participants provided by authors.

[4] Data from all histology cases included, not just NSCLC cases as displayed in paper.

[5] Texas discovery and replication samples excluded as analysed previously in Amos et al. (2008). Never-smoking lung cancer cases and controls also excluded.

2.3.2 Characteristics of included studies

A total of 37 studies published between 2006 and 2010 were identified for inclusion in the meta-analysis. Of these, 19 studies (comprising $k = 57$ independent samples, and a further $k = 15$ duplicate samples) provided data contributing to the meta-analysis (Amos et al., 2008; Breitling et al., 2009; Broderick et al., 2009; X. Chen et al., 2009; Etter et al., 2009; Freathy et al., 2009; Greenbaum et al., 2006; Greenbaum, Rigbi, Teltsh, & Lerer, 2009; Grucza et al., 2008; Keskitalo et al., 2009; Lambrechts et al., 2010; Landi et al., 2009; Le Marchand et al., 2008; Lips et al., 2009; Schwartz et al., 2009; Shiraishi et al., 2009; Spitz et al., 2008; Young et al., 2008; Zienolddiny et al., 2009). The remaining 18 studies identified for inclusion did not contribute data, as data from these studies were not available or the sample(s) featured had been included in another study which we had already included in our analyses (refer to Tables 2.1 and 2.2) (Baker et al., 2009; Caporaso et al., 2009; Conti et al., 2008; Tobacco and Genetics Consortium, 2010; Hung et al., 2008; Liu et al., 2010; Liu et al., 2008; McKay et al., 2008; Pillai et al., 2009; Ray et al., 2010; Rigbi et al., 2008; Sherva et al., 2008; Thorgeirsson et al., 2008; Thorgeirsson et al., 2010; Weiss et al., 2008; Wu et al., 2009; Yang et al., 2010; Young et al., 2009).

A total of 50 samples provided data on participants of predominantly European ancestry, and seven on participants of other ancestry. Twenty-one samples reported data on control/population samples, and 36 on disease/partial disease samples (e.g., lung cancer cases). Forty-four samples reported data on rs1051730, and 27 on rs16969968 (NB: $k = 15$ samples reported data on both SNPs). Two samples reported genotype frequencies that deviated substantially from HWE (NB: one additional non-HWE sample was excluded from analyses as the homozygous risk genotype group contained only one participant). Minor allele frequencies of

rs16969968 (A) and rs1051730 (T) ranged from 0.03-0.43 (median = 0.35). The wide ranges were primarily driven by the inclusion of non-European samples in which the minor alleles were rare.

2.3.3 Smoking quantity

Primary analyses. Meta-analysis indicated strong evidence of association between the rs1051730/rs16966968 variants and daily cigarette consumption (Fixed-effects: B = 0.91, *95% CI* 0.77, 1.06, p <0.001; Random effects: B = 1.01, *95% CI* 0.81, 1.22, p <0.001) (Table 2.3, Figure 2.2). There was evidence of moderate between-study heterogeneity (Q [56] = 85.46, p = 0.007, I^2 = 34%). Regression coefficient units refer to increases/decreases in daily cigarette consumption per copy of the minor allele(s).

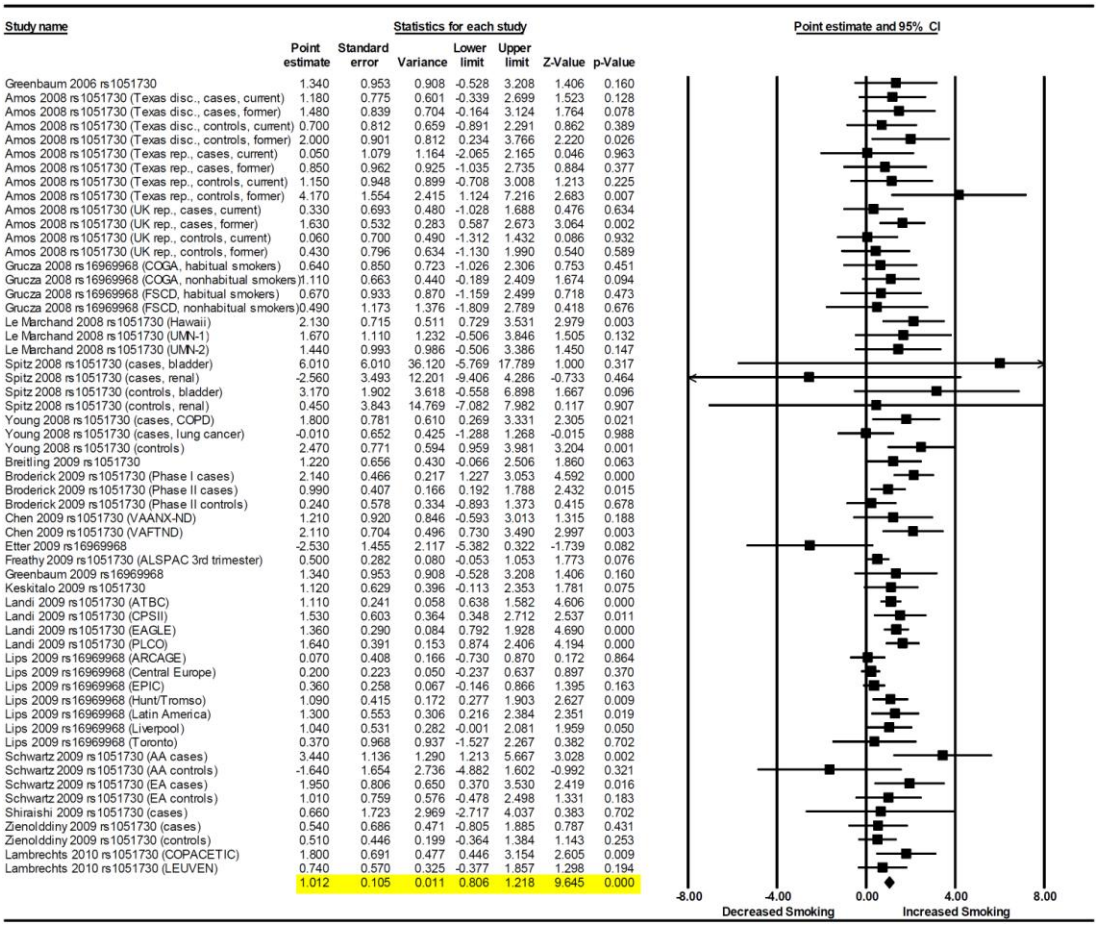
Egger's test indicated weak evidence of small study bias (t [55] = 1.84, $p_{one\text{-}tailed}$ = 0.036). This is visually presented in a funnel plot in Figure 2.3. To adjust for this we utilised Duval and Tweedie's 'trim and fill' method (Duval & Tweedie, 2000). This method removes studies with outlying effect size values identified on funnel plots until symmetry is achieved, and then replaces these along with imputed "mirror" values in order to retain symmetry. This correction had minimal effect on the overall effect estimate (adjusted value: B = 0.85, *95% CI* 0.62, 1.07). There was no evidence of an association between effect size estimate and year of publication (B = -0.17, *95% CI* -0.47, 0.13, p = 0.27).

Table 2.3. Meta-analysis of rs1051730/rs16966968 and heaviness of smoking: Full and stratified analyses.

		Fixed effects						Random effects				
	k	Effect size	95% CI		*p*-value	I^2	p_{diff}	Effect size	95% CI		*p*-value	p_{diff}
Full model	57	0.915	0.769	1.060	<0.001	34%	NA	1.012	0.806	1.218	<0.001	NA
HWE												
Yes	55	0.898	0.752	1.045	<0.001	34%	0.089	0.989	0.781	1.196	<0.001	0.12
No	2	1.894	0.757	3.032	0.001	0%		1.894	0.757	3.032	0.001	
Ancestry												
European	50	0.887	0.739	1.036	<0.001	34%	0.059	0.971	0.763	1.178	<0.001	0.18
Other	7	1.634	0.874	2.394	<0.001	20%		1.584	0.712	2.456	<0.001	
Disease state												
No	21	0.838	0.533	1.143	<0.001	33%	0.57	0.968	0.547	1.390	<0.001	0.79
Yes/Partial	36	0.937	0.772	1.103	<0.001	37%		1.035	0.798	1.273	<0.001	
SNP												
rs1051730	44	1.144	0.964	1.323	<0.001	19%	<0.001	1.170	0.952	1.388	<0.001	0.028
rs16969968	27	0.648	0.444	0.852	<0.001	30%		0.775	0.499	1.051	<0.001	

Results under an additive model of genetic action displayed.

Figure 2.2. Meta-analysis of association of rs1051730/rs16969968 with heaviness of smoking.

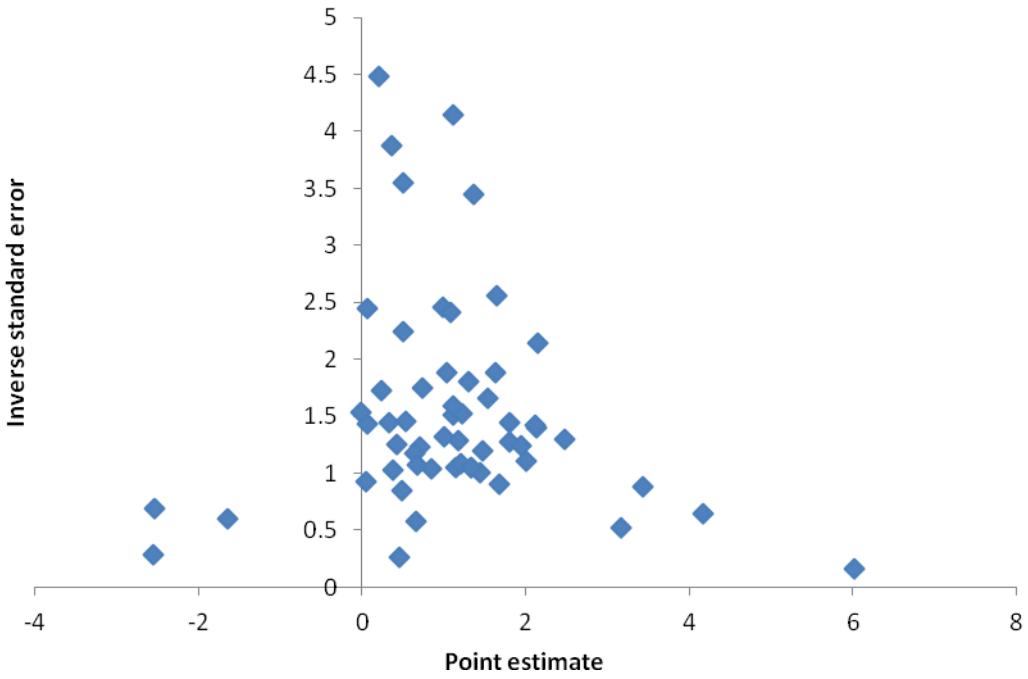


Study name	Point estimate	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Greenbaum 2006 rs1051730	1.340	0.953	0.908	-0.528	3.208	1.406	0.160
Amos 2008 rs1051730 (Texas disc., cases, current)	1.180	0.775	0.601	-0.339	2.699	1.523	0.128
Amos 2008 rs1051730 (Texas disc., cases, former)	1.480	0.839	0.704	-0.164	3.124	1.764	0.078
Amos 2008 rs1051730 (Texas disc., controls, current)	0.700	0.812	0.659	-0.891	2.291	0.862	0.389
Amos 2008 rs1051730 (Texas disc., controls, former)	2.000	0.901	0.812	0.234	3.766	2.220	0.026
Amos 2008 rs1051730 (Texas rep., cases, current)	0.050	1.079	1.164	-2.065	2.165	0.046	0.963
Amos 2008 rs1051730 (Texas rep., cases, former)	0.850	0.962	0.925	-1.035	2.735	0.884	0.377
Amos 2008 rs1051730 (Texas rep., controls, current)	1.150	0.948	0.899	-0.708	3.008	1.213	0.225
Amos 2008 rs1051730 (Texas rep., controls, former)	4.170	1.554	2.415	1.124	7.216	2.683	0.007
Amos 2008 rs1051730 (UK rep., cases, current)	0.330	0.693	0.480	-1.028	1.688	0.476	0.634
Amos 2008 rs1051730 (UK rep., cases, former)	1.630	0.532	0.283	0.587	2.673	3.064	0.002
Amos 2008 rs1051730 (UK rep., controls, current)	0.060	0.700	0.490	-1.312	1.432	0.086	0.932
Amos 2008 rs1051730 (UK rep., controls, former)	0.430	0.796	0.634	-1.130	1.990	0.540	0.589
Grucza 2008 rs16969968 (COGA, habitual smokers)	0.640	0.850	0.723	-1.026	2.306	0.753	0.451
Grucza 2008 rs16969968 (COGA, nonhabitual smokers)	1.110	0.663	0.440	-0.189	2.409	1.674	0.094
Grucza 2008 rs16969968 (FSCD, habitual smokers)	0.670	0.933	0.870	-1.159	2.499	0.718	0.473
Grucza 2008 rs16969968 (FSCD, nonhabitual smokers)	0.490	1.173	1.376	-1.809	2.789	0.418	0.676
Le Marchand 2008 rs1051730 (Hawaii)	2.130	0.715	0.511	0.729	3.531	2.979	0.003
Le Marchand 2008 rs1051730 (UMN-1)	1.670	1.110	1.232	-0.506	3.846	1.505	0.132
Le Marchand 2008 rs1051730 (UMN-2)	1.440	0.993	0.986	-0.506	3.386	1.450	0.147
Spitz 2008 rs1051730 (cases, bladder)	6.010	6.010	36.120	-5.769	17.789	1.000	0.317
Spitz 2008 rs1051730 (cases, renal)	-2.560	3.493	12.201	-9.406	4.286	-0.733	0.464
Spitz 2008 rs1051730 (controls, bladder)	3.170	1.902	3.618	-0.558	6.898	1.667	0.096
Spitz 2008 rs1051730 (controls, renal)	0.450	3.843	14.769	-7.082	7.982	0.117	0.907
Young 2008 rs1051730 (cases, COPD)	1.800	0.781	0.610	0.269	3.331	2.305	0.021
Young 2008 rs1051730 (cases, lung cancer)	-0.010	0.652	0.425	-1.288	1.268	-0.015	0.988
Young 2008 rs1051730 (controls)	2.470	0.771	0.594	0.959	3.981	3.204	0.001
Breitling 2009 rs1051730	1.220	0.656	0.430	-0.066	2.506	1.860	0.063
Broderick 2009 rs1051730 (Phase I cases)	2.140	0.466	0.217	1.227	3.053	4.592	0.000
Broderick 2009 rs1051730 (Phase II cases)	0.990	0.407	0.166	0.192	1.788	2.432	0.015
Broderick 2009 rs1051730 (Phase II controls)	0.240	0.578	0.334	-0.893	1.373	0.415	0.678
Chen 2009 rs1051730 (VAANX-ND)	1.210	0.920	0.846	-0.593	3.013	1.315	0.188
Chen 2009 rs1051730 (VAFTND)	2.110	0.704	0.496	0.730	3.490	2.997	0.003
Etter 2009 rs16969968	-2.530	1.455	2.117	-5.382	0.322	-1.739	0.082
Freathy 2009 rs1051730 (ALSPAC 3rd trimester)	0.500	0.282	0.080	-0.053	1.053	1.773	0.076
Greenbaum 2009 rs16969968	1.340	0.953	0.908	-0.528	3.208	1.406	0.160
Keskitalo 2009 rs1051730	1.120	0.629	0.396	-0.113	2.353	1.781	0.075
Landi 2009 rs1051730 (ATBC)	1.110	0.241	0.058	0.638	1.582	4.606	0.000
Landi 2009 rs1051730 (CPSII)	1.530	0.603	0.364	0.348	2.712	2.537	0.011
Landi 2009 rs1051730 (EAGLE)	1.360	0.290	0.084	0.792	1.928	4.690	0.000
Landi 2009 rs1051730 (PLCO)	1.640	0.391	0.153	0.874	2.406	4.194	0.000
Lips 2009 rs16969968 (ARCAGE)	0.070	0.408	0.166	-0.730	0.870	0.172	0.864
Lips 2009 rs16969968 (Central Europe)	0.200	0.223	0.050	-0.237	0.637	0.897	0.370
Lips 2009 rs16969968 (EPIC)	0.360	0.258	0.067	-0.146	0.866	1.395	0.163
Lips 2009 rs16969968 (Hunt/Tromso)	1.090	0.415	0.172	0.277	1.903	2.627	0.009
Lips 2009 rs16969968 (Latin America)	1.300	0.553	0.306	0.216	2.384	2.351	0.019
Lips 2009 rs16969968 (Liverpool)	1.040	0.531	0.282	-0.001	2.081	1.959	0.050
Lips 2009 rs16969968 (Toronto)	0.370	0.968	0.937	-1.527	2.267	0.382	0.702
Schwartz 2009 rs1051730 (AA cases)	3.440	1.136	1.290	1.213	5.667	3.028	0.002
Schwartz 2009 rs1051730 (AA controls)	-1.640	1.654	2.736	-4.882	1.602	-0.992	0.321
Schwartz 2009 rs1051730 (EA cases)	1.950	0.806	0.650	0.370	3.530	2.419	0.016
Schwartz 2009 rs1051730 (EA controls)	1.010	0.759	0.576	-0.478	2.498	1.331	0.183
Shiraishi 2009 rs1051730 (cases)	0.660	1.723	2.969	-2.717	4.037	0.383	0.702
Zienolddiny 2009 rs1051730 (cases)	0.540	0.686	0.471	-0.805	1.885	0.787	0.431
Zienolddiny 2009 rs1051730 (controls)	0.510	0.446	0.199	-0.364	1.384	1.143	0.253
Lambrechts 2010 rs1051730 (COPACETIC)	1.800	0.691	0.477	0.446	3.154	2.605	0.009
Lambrechts 2010 rs1051730 (LEUVEN)	0.740	0.570	0.325	-0.377	1.857	1.298	0.194
	1.012	0.105	0.011	0.806	1.218	9.645	0.000

Results under an additive model of genetic action displayed. Data analysed within a random-effects framework. Results from individual studies are listed in order according to year of publication.

Individual estimates weighted using Der Simonian and Laird methods. Overall effect size: $B = 1.01$, *95% CI* 0.81, 1.22, $p<0.001$. Point estimates refer to increases/decreases in daily cigarette consumption, units representing whole cigarettes.

Figure 2.3. Funnel plot illustrating evidence of small study bias in pooled meta-analysis.



Stratified analyses. Results from all stratified analyses, under both fixed- and random-effects models, are displayed in Table 2.3. Random-effects model outcomes are presented here. Evidence for an association between rs1051730/rs16969968 variants and heaviness of smoking was observed irrespective of stratification by study level characteristics.

Two of the 57 samples included in our analysis deviated from HWE. We compared this pair of samples to the group of 55 samples which did not deviate from HWE. In both groups there was strong evidence to suggest an association between rs1051730/rs16969968 and daily cigarette consumption (HWE: *B* = 0.99, *95% CI* 0.78, 1.20, *p* <0.001; Non-HWE: *B* = 1.89, *95% CI* 0.76, 3.03, *p* = 0.001).There was no clear evidence to suggest a difference in effect size estimates between groups (p_{diff} = 0.12).

There was strong evidence of an association between rs1051730/rs16969968 and daily cigarette consumption in both European and Other groups (European: *B* = 0.97, *95% CI* 0.76, 1.18, *p* <0.001; Other: *B* = 1.58, *95% CI* 0.71, 2.46, *p* <0.001).There was no clear evidence to suggest that this effect size differed between groups (p_{diff} = 0.18).

There was strong evidence of an association between rs1051730/rs16969968 and daily cigarette consumption in both the control/population group and the disease/partial disease group (control: *B* = 0.97, *95% CI* 0.55, 1.39, *p* <0.001; disease/partial: *B* = 1.04, *95% CI* 0.80, 1.27, *p* <0.001). There was no evidence for a difference in effect size estimates between groups (p_{diff} = 0.79).

There was strong evidence of an association between both rs1051730 and rs16969968 SNPs and daily cigarette consumption (rs1051730: *B* = 1.17, *95% CI* 0.95, 1.39, *p* <0.001; rs16969968: *B* = 0.77, *95% CI* 0.50, 1.05, *p* <0.001), and this

effect size appeared to differ between groups (p = 0.028). However, although this difference was qualitatively observed in the sub-set of samples (k = 14) which contained data on both SNPs, with a slightly larger effect size observed for rs1051730 (B = 1.09, *95% CI* 0.72, 1.46, p <0.001) compared with rs16966968 (B = 1.05, *95% CI* 0.65, 1.46, p <0.001), this difference did not achieve statistical significance (p_{diff} = 0.89), suggesting that the observed difference in the full meta-analysis may be due to confounding arising from other study- or sample-level differences. One sample reporting data on both SNPs was excluded from analyses as the homozygous risk genotype group contained only one participant.

Egger's test indicated no evidence of small study bias for SNP rs1051730 (t [42] = 0.92, $p_{one\text{-}tailed}$ = 0.18). Evidence of small study bias was observed for SNP rs16969968 however (t [25] = 2.01, $p_{one\text{-}tailed}$ = 0.028). We utilised Duval and Tweedie's 'trim and fill' method to adjust for this, which led to a reduction in the overall effect estimate for this SNP (adjusted value: B = 0.49, *95% CI* 0.19, 0.79).

2.4 Discussion

Our data suggest compelling evidence for a small effect of the rs16969968/rs1051730 SNPs on daily cigarette consumption, equivalent to a per allele effect of approximately one cigarette per day. Interestingly, rs1051730 may provide a stronger signal than rs16969968, although evidence for this is indirect and should therefore be treated with caution. No evidence for a difference in effect size between groups was observed in other stratified analyses (i.e., ancestry, disease state). Strong evidence for an association between rs16969968/rs1051730 SNPs and daily cigarette consumption was observed irrespective of study level characteristics, suggesting that the association is robust.

The nicotinic acetylcholine receptor, to which nicotine binds, is a plausible and biologically relevant candidate for smoking aetiology, as previously discussed (Chapter 1). Although the mis-sense mutation rs16969968 appears to be of functional significance, a larger effect was observed for the synonymous SNP rs1051730 in full stratified analyses. Whilst this variant appears to be of no functional significance, which to some extent limits the interpretation of the observed association, as a single marker it appears superior to rs16969968 with regards to determining variation in smoking quantity. It is also noteworthy that evidence of small study bias was observed for rs16969968 but not for rs1051730. Adjusting for this bias increased the difference in effect estimate between variants. It is possible that rs1051730 is a strong tagging SNP for functional haplotypes in this region and it would therefore be important to focus research efforts on identifying these. It is crucial to note however that the difference in effect size estimates between variants was only qualitatively observed in the subset of samples which contained data on both rs1051730 and rs16969968. This may be due to the limited number of studies examined (k = 14) or driven by other study- or sample-level differences. A large scale study directly comparing both variants would be required to answer this question definitively.

The primary limitation of our meta-analysis was that we did not have the data necessary to perform a joint SNP analysis in which the effects of one variant were conditioned on the other. This analysis would have enabled us to comment more authoritatively on the difference in genetic signal between these two SNPs, if any, which are known to be in linkage disequilibrium (LD). Linkage disequilibrium refers to the non-random association of alleles at different loci, and may be inferred from genotype frequency data. The strength of association between loci may be assessed using different measures, including r^2 and D'. These measures range from 0 (no

disequilibrium) to 1 (complete disequilibrium) (Wall & Pritchard, 2003). The strong LD observed between the variants examined here (European: $r^2 = 0.902$, Japanese/Chinese: $r^2 = 1.000$, African: r^2 = unavailable; calculated using HapMap data in conjunction with SNAP [http://www.broadinstitute.org/mpg/snap/ldsearchpw.php]) is certainly likely to have impacted on our ability to detect differences in effect estimates between SNPs. A large-scale study would certainly be required to detect potential differences. An additional limitation of our meta-analysis was that the procedures used allowed only comparable data to be combined. As such a number of studies that would have ideally been included in our analysis had to be excluded, such as those examining extreme smoking quantity phenotypes (e.g., Stevens et al., 2008). We were also unable to include data from studies reporting only categorical smoking quantity data by genotype (e.g., Thorgeirsson et al., 2008). An additional shortcoming was that we were only able to investigate a limited number of study-level characteristics. It is of note, however, that the analysis of study-level characteristics is indirect, and may lead to ecological fallacy. Any differences observed should be considered hypothesis-generating to be followed-up in appropriately designed primary studies. Additionally, it is of note that methods employed to correct for publication bias, such as Duval & Tweedie's 'trim and fill' approach as utilised here, are not widely accepted and rest on certain assumptions (see Munafo, Clark, & Flint, 2004). As such, corrected findings should be interpreted with caution. Finally, a comprehensive assessment of study quality and potential bias was not considered in this meta-analysis. As Little et al. (2009) highlight, the manner in which a study is conducted and choices made regarding study design and data analysis may all potentially influence the magnitude and direction of results in genetic association studies.

Potential methods for incorporating information on study quality into meta-analyses include quality weighting (Rosenthal, 1991) and subgroup analysis by quality. Whilst we did not assess and consider all aspects of study quality and potential sources of bias, we did conduct subgroup analyses contrasting samples which did and did not deviate from HWE. Deviations from HWE are typically indicative of genotyping error which can impact study power. No difference in effect size was observed between these two groups however.

In conclusion, our analyses confirm that two variants (rs16969968 and rs1051730) located in the nicotinic acetylcholine receptor gene cluster *CHRNA5-A3-B4* are robustly associated with heaviness of smoking. Interestingly, rs1051730 *may* provide a stronger signal than rs16969968, although evidence for this is indirect. Much variability in this phenotype remains to be determined however. Smoking is a complex behaviour determined by both genetic and environmental factors. It is likely that many other loci will contribute to this phenotype, as will multiple environmental factors. It is also important to also consider gene-gene interactions (epistasis), as well as gene-environment interactions, whereby the expression of a genetic effect is modified by environmental exposure(s). It is of note that genetic influences on smoking phenotypes such as heaviness of smoking and nicotine dependence can only be expressed following initial exposure to tobacco. Therefore, environmental factors which influence likelihood of initial exposure are of particular interest, as they hold the potential to modify expression of genetic effects. This is explored further in the following chapter.

Chapter 3

Genetic Epidemiology:

The Association between *CHRNA3*, *BDNF*, Parental Monitoring and Smoking Initiation: A Longitudinal Study

3.1 Introduction

3.1.1 The Avon Longitudinal Study of Parents and Children

The Avon Longitudinal Study of Parents and Children (ALSPAC) (http://www.alspac.bris.ac.uk) has been an essential resource for the completion of my doctoral studies. Both the current and subsequent experimental chapters of this thesis are based on data gathered from (or plans to gather from) ALSPAC participants. Within this section we discuss ALSPAC in detail, and highlight why it is particularly well-suited for the study of behavioural genetics.

The Avon Longitudinal Study of Parents and Children, also known as 'Children of the 90s', is a transgenerational prospective observational study (Boyd et al., 2012). It was founded to establish how genetic and environmental factors influence health and development (Fraser et al., 2012). In brief, 14,541 pregnant women resident in the former county of Avon, UK, with expected delivery dates between 1st April 1991 and 31st December 1992 were recruited for this study. At the planning stages of ALSPAC, the Avon population was considered similar to that of the whole of Great Britain (Golding, Pembrey, & Jones, 2001). Phenotypic and environmental data, alongside genetic information and biological samples, have been collected from mothers and their offspring at multiple time points throughout the course of the study (Boyd et al., 2012). The breadth and frequency of data collection,

including the availability of repeat measures, is a particular strength of the study, and of particular importance to the reported scheme of work.

> *"The ALSPAC resource has a scale and richness that is unprecedented in epidemiological studies"*
>
> (Boyd et al., 2012, p9)

Data on the children have been collected at 68 time points between birth and 18 years of age. This has included 34 child-completed questionnaires, nine clinical assessments ("Focus" clinics), and 25 questionnaires about the child completed by mothers/primary caregivers (Boyd et al., 2012). Data on smoking behaviour (as reported by the child) has been collected in brief at the ages of eight and 10 years (clinic setting), and in more detail at the ages of 12, 13, 14, 15, 16 and 17 years (assessed in a clinic setting or reported via postal questionnaire).

Data on the mothers has been ascertained through 18 self-report questionnaires (from pregnancy to 20 years postnatal), medical records, opportunistic clinic assessments, and a recent clinic assessment focused specifically on mothers (FoM1) (Fraser et al., 2012). A wealth of data has been collected on maternal smoking behaviour at multiple time points throughout the study. This includes information on smoking during pregnancy.

DNA samples are available for 11,343 children and 10,321 mothers. Genome-wide data are available for 8,365 children and over 10,000 mothers, and complete genome sequencing data will soon be available for 2,000 children. DNA methylation data for 1,000 mother and child pairs are also pending (Boyd et al., 2012; Fraser et al., 2012).

The large sample size, richness of phenotypic and environmental data, combined with the availability of genetic information, make ALSPAC a particularly

well-suited resource for the study of behavioural genetics, including the field of molecular genetics.

3.1.2 CHRNA3, BDNF, parental monitoring and smoking initiation

An estimated 82,000 to 99,000 young people worldwide start smoking every day (Lando et al., 2010). In order to develop appropriate preventative measures, it is important to identify the causes of smoking initiation, and also factors underlying progression from first use to dependence.

In a recent meta-analysis of genome-wide association studies of smoking behaviours, a nonsynonymous single nucleotide polymorphism (SNP) rs6265 in *BDNF,* which encodes brain-derived neurotrophic factor, was found to be most strongly associated with smoking initiation, determined through comparison of ever versus never smokers (Furberg et al., 2010). BDNF plays an important role in the survival and differentiation of dopaminergic neurons, and has previously been linked to substance use (Gratacos et al., 2007). Another genetic variant that has proven of particular importance in relation to smoking-related phenotypes is rs1051730 in *CHRNA3* (as discussed previously). Although an association between this locus and smoking initiation has not been observed, it has previously been associated with positive initial smoking experiences (Sherva et al., 2008), and has been consistently associated with other smoking-related phenotypes.

Genetic influences on smoking phenotypes such as nicotine dependence can only be expressed following initial exposure to tobacco. Therefore, environmental factors which influence likelihood of initial exposure may modify the expression of genetic effects. One such factor of *a priori* relevance is parental monitoring. A substantial body of evidence has demonstrated an association between level of parental monitoring and adolescent substance use, including smoking (Bohnert, Rios-

Bedoya, & Breslau, 2009; Chilcoat, Dishion, & Anthony, 1995; Piko & Kovacs, 2010; Pokhrel, Unger, Wagner, Ritt-Olson, & Sussman, 2008). The relationship consistently demonstrated between level of monitoring and smoking initiation may be explained as follows: Adolescents who are poorly monitored by parents/guardians are more likely to be regularly exposed to 'risky' environments where cigarettes are available, and/or to smoking peers, thus increasing risk of smoking initiation, whilst the converse is true for closely monitored children.

In addition to its well-established, independent effect on smoking behaviour, parental monitoring has also been reported to modify genetic risk for smoking. In a study of Finnish twins, Dick et al. (2007) showed that the relative importance of genetic influences on smoking behaviour changed substantially as a function of parental monitoring, genetic effects significantly decreasing in importance as parental monitoring levels increased. The primary interpretation drawn from this study was that a background of low parental monitoring created an environment allowing for greater opportunity to express genetic predispositions. Building upon these findings, researchers have now begun to explore the impact of environmental factors on associations between specific genes and smoking-related phenotypes. Chen et al. (2009), for example, recently examined whether level of parental monitoring during early adolescence modified the risk of nicotine dependence associated with rs16969968 in a US-based community sample. Both parental monitoring and rs16969968 were associated with risk of nicotine dependence. In addition, and of key interest here, expression of the genetic effect was modified by parental monitoring level; risk for nicotine dependence significantly increased with the risk genotype of rs16969968 when combined with the lowest quartile of parental monitoring. It is of note that the study of gene × environment interactions,

exemplified by Caspi et al. (2003), has come under substantial criticism in recent years (e.g., Flint & Munafo, 2008; Munafo & Flint, 2009). This approach was borne from the notion that such interactions may "illuminate the aetiology and genetic architecture of behavioural phenotypes" (Munafo & Flint, 2009). However, inconsistencies and persistent failures to replicate effects have dominated this field of research. One likely reason for this is the potential this approach offers for data dredging (i.e., manipulation of multiple parameters), and associated use of multiple (and often unreported) statistical tests. Such techniques may be utilised to identify 'significant', and thus publishable, findings, yet serve only to confuse the existent literature. It is important to note at this juncture that in this study we have a strong *a priori* reason for expecting the effect of interest to be restricted to a specific subgroup. Our primary variant of interest, rs1051730, is located in a gene responsible for encoding the α_3 nicotinic receptor subunit, thus expression of its effect may only be expressed following initial exposure to tobacco. As discussed, adolescents who are poorly monitored are likely to experience opportunities to smoke, and smoke regularly, thus enabling expression of said genetic effect.

Smoking initiation, considered as a binary outcome, has limitations with respect to understanding the complex processes involved in smoking acquisition. Whilst many adolescents will try a cigarette (i.e., very first use), not all will progress through experimentation to regular use. Heritability estimates differ for initiation relative to progression of smoking (Fowler et al., 2007), suggestive of differing underlying causes. These issues have been discussed at length elsewhere (National Cancer Institute, 2009). It is therefore important to consider initial smoking trajectories. By capturing the complexities of initial cigarette use, these may prove to be more informative phenotypes for examination in genetic association studies.

Using data from a prospectively assessed cohort, we sought to determine the association between rs6265 (*BDNF*) and rs1051730 (*CHRNA3*) and smoking initiation. The variant rs6265 was included to serve as a positive control given its previously established association with initiation. Initiation was assessed both in terms of 'ever use' of cigarettes, and also in terms of initial smoking trajectories – a novel, sophisticated phenotype which captures the complexities of initial cigarette use, determined using repeated measures of smoking frequency. Given that rs1051730 is located in a gene responsible for encoding the α_3 nicotinic receptor subunit, it was hypothesised that its effects would not be associated with ever (i.e., very first) use of cigarettes, but may be associated with initial smoking trajectories, namely those capturing progression from initial exposure to tobacco (engaging nicotinic receptors) through to regular use. We also sought to determine potential modification in the expression of these genetic effects by a well-established environmental risk factor for smoking initiation, namely parental monitoring. Given that this variable may plausibly influence the probability of exposure to tobacco/opportunities to smoke, I hypothesised that level of parental monitoring would moderate the expression of the predicted genetic effect.

3.2 Methods

3.2.1 Participants

All participants were drawn from ALSPAC (Boyd et al., 2012). From an initial 14,541 pregnancies, with a total of 14,676 foetuses, 14,062 live births took place. For reasons of confidentiality, data on the 13 triplet and quadruplet children were not available for analysis. Our starting sample consisted of the 13,976 singletons and twins who survived until one year of age. Ethics approval for this

study was obtained from the ALSPAC Ethics and Law Committee, which is registered as an Institutional Review Board.

3.2.2 Measurements

Smoking initiation. Smoking 'initiation' was assessed using two distinct measures: 1. Ever use of cigarettes; and 2. Initial trajectories of smoking behaviour. Ever use data was collected at age 16 years via a postal questionnaire. This was the most recent wave of data available at time of analysis. The specific question posed was as follows: "Have you ever smoked a cigarette (including roll-ups)?", to which participants responded "yes" or "no". Derivation of initial smoking trajectory categories is described in full in Heron et al. (2011). This approach was adopted to examine smoking trajectory phenotypes in the current study. Briefly, latent class analysis was used to analyse repeated measures of smoking frequency in the ALSPAC cohort (data gathered at ages 14, 15 and 16 years), enabling the identification of distinct smoking behaviour profiles. Smoking frequency was defined as a four-category ordinal variable comprised of the following categories: "none"; "less than weekly"; "weekly"; "daily". Four distinct initial trajectories of smoking behaviour were determined: non-smokers, experimenters, late-onset regular smokers, and early-onset regular smokers. Each individual in our sample was assigned a probability of belonging to each of these four categories based on repeated measures of smoking behaviour collected at ages 14, 15, and 16 years. The data gathered at ages 14 and 16 years were collected via postal questionnaire, whilst the data gathered at age 15 years were collected in a clinic setting via a computer terminal. Multiple imputation was used to enable classification of individuals with partially missing data to a particular trajectory, thus enabling us to maximise sample size (see Figure 3.2).

Genetic data. DNA was extracted as described previously (Jones et al., 2000). Genotyping was undertaken by KBioscience Ltd. (www.kbioscience.co.uk), using a proprietary competitive allele specific PCR system (KASPar) for SNP analysis. For this study we focused on two SNPs, namely rs6265 (*BDNF*) and rs1051730 (*CHRNA3*).

Parental monitoring. Level of parental monitoring was assessed in a clinic setting. It was assessed at age 11 years, thus preceding our assessment of smoking behaviour in the children. Parents/guardians were not present. During a short, structured interview the children were asked three questions pertaining to parental monitoring: 1. Whether the grown-ups that they live with knew all the other children that the child 'hangs around' with (responses: yes, all; quite a lot; a few; none); 2. Whether grown-ups knew about what the child did with other children (responses: everything; most things; few things; nothing); and 3. Whether the child would ask a grown up for help if they were having problems with their friends or other children (responses: yes; most of the time; occasionally; definitely not). Responses were made on a four-point scale, scored zero to three. The participant's responses to these three items were summed to form a total score ranging between zero and nine, higher scores indicating lower levels of parental monitoring. Given the uneven distribution of scores (approximately 85% of individuals scored between zero and three in both samples), application of pre-specified cut-points based on score were prohibited due to low cell counts. As such, post-hoc determination of cut-points was required. Scores of zero, one, two and three were grouped into category '0' (~85% of samples), and scores ranging from four to nine into category '1' (~15% of samples).. Given this somewhat unorthodox approach to category classification, sensitivity

analyses examining the effect of altering this cut-off were also conducted (see Section *3.3.5*).

3.2.3 Statistical analyses

A series of chi-square tests were first conducted to explore associations between genetic variants, parental monitoring, and each of the smoking initiation phenotypes. Several additional variables previously identified as risk factors for initial smoking trajectories were also analysed, including: sex, housing tenure (coded as owned/mortgaged, privately rented, subsidised housing), crowding status (coded as the ratio of number of residents to number of rooms in house), maternal education attainment (coded as no high school qualifications, high school, beyond high school), maternal smoking during pregnancy (first trimester), and parity (coded as whether study child was first, second, third child or greater) (see Heron et al., 2011).

Logistic regression was used to test main effects of both variants and parental monitoring on smoking initiation (ever use), and to investigate gene-environment interactions. Multinomial logistic regression was used for initial smoking trajectory analyses. Additional regression analyses were conducted including the covariates outlined above, which had been selected for inclusion in regression models *a priori* given their previous identification as risk factors for smoking initiation. In all instances, likelihood ratio tests were utilised to compare log likelihood values between statistical models. Genetic variants were considered as linear variables for the purposes of analysis, coded in terms of frequency of the minor allele (i.e., 0, 1 or 2). A χ^2 test was used to assess whether or not the genotype frequencies of both SNPs were in Hardy-Weinberg equilibrium.

All statistical analyses were conducted using Stata version 11.1.

3.3 Results

3.3.1 Sample derivation

The initial sample for both sets of analyses consisted of the 13,976 singletons and twins in the ALSPAC cohort who survived until one year of age. However, our final samples were considerable smaller (see Figures 3.1 and 3.2), due to sample attrition. We further restricted analyses to individuals with the required data on smoking behaviour (at ages 14, 15 and/or 16 years), parental monitoring (at age 11 years) and genetics. Individuals who had reported ever smoking at 10 years (i.e., prior to the parental monitoring assessment) were excluded, and our final sample was limited to individuals of European ancestry. Our final samples thus consisted of 2,687 and 3,771 individuals for ever use and smoking trajectory analyses respectively. Figures 3.1 and 3.2 illustrate participant flow.

Figure 3.1. Ever use sample – participant flow.

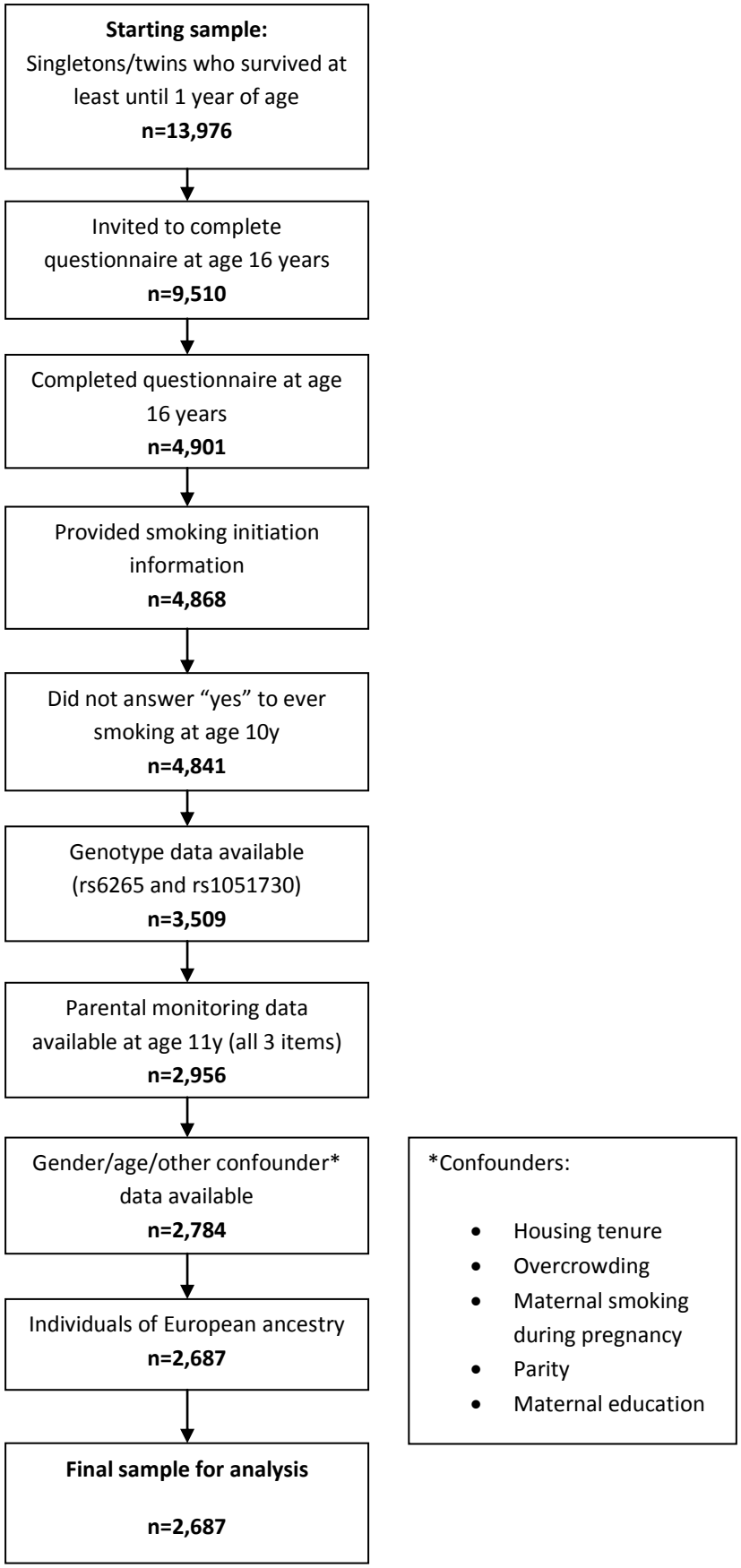
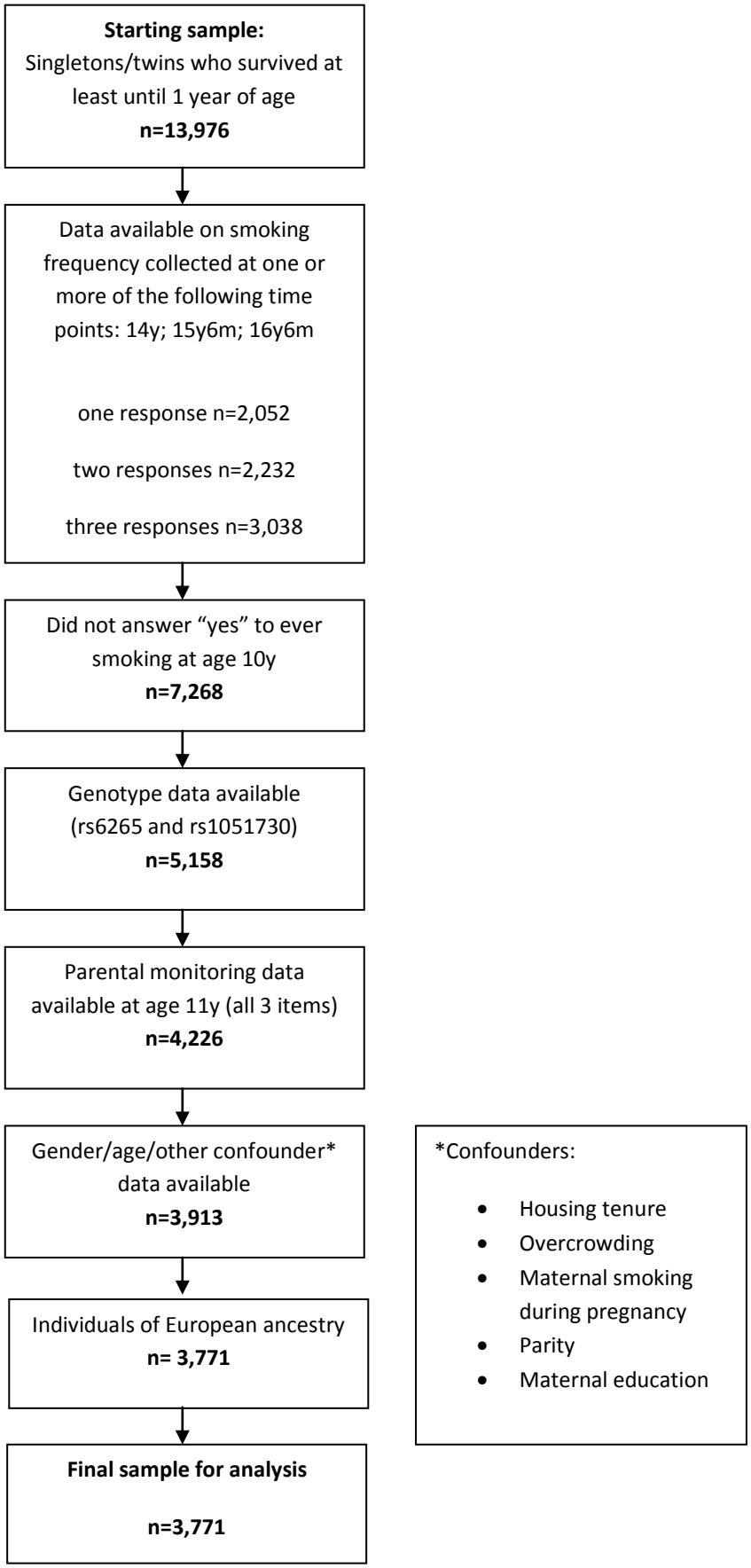


Figure 3.2. Smoking trajectories sample – participant flow.



3.3.2 Characteristics of participants

Ever use sample. Approximately one half of the sample reported having ever smoked a cigarette by the age of 16 (see Table 3.1 for sample characteristics), with girls being more likely than boys to have tried smoking by this age ($p < 0.001$). Other factors influencing ever use of cigarettes by this age included overcrowding in the home ($p = 0.006$), maternal smoking during pregnancy ($p < 0.001$), and parity ($p < 0.001$). Genotype frequencies for rs1051730 were: C:C 45.4% C:T 43.0% T:T 11.6%, and for rs6265 were: G:G 66.8% G:A 29.4% A:A 3.8%. These genotype frequencies were in Hardy-Weinberg equilibrium (rs1051730: $p = 0.13$; rs6265: $p = 0.17$).

Table 3.1. Sample characteristics for ever use of cigarettes.

	Total		Ever smoked by 16y		Never smoked by 16y		χ^2 test *p* value
	n	%	n	%	n	%	
Total	2687	100%	1278	47.6%	1409	52.4%	
Gender							
Male	1137	42.3%	426	33.3%	711	50.5%	<0.001
Female	1550	57.7%	852	66.7%	698	49.5%	
SNP rs1051730							
CC	1221	45.4%	599	46.9%	622	44.1%	0.35
CT	1155	43.0%	537	42.0%	618	43.9%	
TT	311	11.6%	142	11.1%	169	12.0%	
SNP rs6265							
GG	1794	66.8%	840	65.7%	954	67.7%	0.16
GA	790	29.4%	395	30.9%	395	28.0%	
AA	103	3.8%	43	3.4%	60	4.3%	
Parental monitoring level							
0 (highest)	2298	85.5%	1058	82.8%	1240	88.0%	<0.001
1 (lowest)	389	14.5%	220	17.2%	169	12.0%	
Housing tenure							
Mortgaged/owned	2389	88.9%	1128	88.3%	1261	89.5%	0.53
Rented	169	6.3%	83	6.5%	86	6.1%	
Subsidised	129	4.8%	67	5.2%	62	4.4%	
Overcrowding							
≤0.5 person/room	1465	54.5%	665	52.0%	800	56.8%	0.006
>0.5 – 0.75 person/room	811	30.2%	389	30.4%	422	30.0%	
>0.75 – 1 person/room	347	12.9%	194	15.2%	153	10.9%	
>1 person/room	64	2.4%	30	2.3%	34	2.4%	
Smoking during pregnancy							
No	2322	86.4%	1065	83.3%	1257	89.2%	<0.001
Yes	365	13.6%	213	16.7%	152	10.8%	
Parity							
First child	1299	48.3%	575	45.0%	724	51.4%	<0.001
Second child	964	35.9%	470	36.8%	494	35.1%	
Third child or higher	424	15.8%	233	18.2%	191	13.6%	
Maternal education							
Qualifications > high school	1363	50.7%	631	49.4%	732	52.0%	0.23
High school qualifications	905	33.7%	433	33.9%	472	33.5%	
No high school qualifications	419	15.6%	214	16.7%	205	14.5%	

Smoking trajectories sample. Of this sample, 81.7% were classified as non-smokers, whilst 8.8% were classified as experimenters, 6.6% as late-onset regular smokers, and 2.8% as early-onset regular smokers (see Table 3.2 for sample characteristics). Exact sample sizes within each trajectory cannot be specified as these values are based on probabilities. Other factors influencing initial smoking trajectories included sex ($p < 0.001$), housing tenure ($p < 0.001$), overcrowding ($p < 0.001$), maternal smoking during pregnancy ($p < 0.001$), parity ($p < 0.001$), and level of maternal education ($p < 0.001$), as previously reported by Heron et al. (2011). Genotype frequencies for rs1051730 were: C:C 45.2% C:T 43.7% T:T 11.1%, and for rs6265 were: G:G 66.9% G:A 29.0% A:A 4.1%. Genotype frequencies for rs1051730 were in Hardy-Weinberg equilibrium ($p = 0.51$), although they were not for rs6265 ($p = 0.008$).

Table 3.2. Sample characteristics for initial smoking trajectories.

	Total		Non-Smoker	Experi-menter	Late-Onset Regular	Early-Onset Regular	χ^2 test *p* value
	n	%	%	%	%	%	
Total	3771	100%	81.7%	8.8%	6.6%	2.8%	
Gender							
Male	1779	47.2%	85.1%	7.3%	5.3%	2.3%	<0.001
Female	1992	52.8%	78.7%	10.2%	7.8%	3.3%	
SNP rs1051730							
CC	1703	45.2%	80.9%	8.7%	7.3%	3.1%	0.19
CT	1649	43.7%	82.0%	9.2%	6.2%	2.7%	
TT	419	11.1%	84.3%	8.2%	5.7%	1.8%	
SNP rs6265							
GG	2524	66.9%	81.5%	8.9%	6.8%	2.8%	0.89
GA	1092	29.0%	82.1%	8.9%	6.3%	2.7%	
AA	155	4.1%	83.2%	7.8%	5.9%	3.1%	
Parental monitoring level							
0 (highest)	3143	83.4%	83.3%	8.5%	5.9%	2.3%	<0.001
1 (lowest)	628	16.7%	73.9%	10.6%	10.3%	5.2%	
Housing tenure							
Mortgaged/owned	3286	87.1%	82.2%	8.9%	6.5%	2.4%	<0.001
Rented	246	6.5%	80.6%	8.9%	7.0%	3.6%	
Subsidised	239	6.3%	75.9%	8.1%	8.4%	7.7%	
Overcrowding							
≤0.5 person/room	1990	52.8%	83.8%	8.8%	5.7%	1.8%	<0.001
>0.5 – 0.75 person/room	1165	30.9%	81.3%	8.1%	7.0%	3.6%	
>0.75 – 1 person/room	507	13.4%	76.5%	10.6%	9.0%	3.9%	
>1 person/room	109	2.9%	73.5%	10.8%	8.2%	7.5%	
Smoking during pregnancy							
No	3209	85.1%	83.3%	8.7%	6.1%	2.0%	<0.001
Yes	562	14.9%	72.7%	9.9%	9.8%	7.6%	
Parity							
First child	1779	47.2%	84.4%	8.1%	5.5%	2.1%	<0.001
Second child	1371	36.4%	80.2%	9.4%	7.5%	3.0%	
Third child or higher	621	16.5%	77.5%	9.8%	8.3%	4.4%	
Maternal education							
Qualifications > high school	1757	46.6%	83.0%	9.4%	5.7%	1.9%	<0.001
High school qualifications	1336	35.4%	81.0%	8.6%	7.3%	3.2%	
No high school qualifications	678	18.0%	79.7%	7.8%	8.0%	4.5%	

3.3.3 Associations of genotype and parental monitoring with smoking behaviour

Logistic regression models of smoking initiation are shown in Tables 3.3 and 3.4. We found strong evidence of an association between parental monitoring and smoking initiation, both for ever use (unadjusted: $p < 0.001$) and smoking trajectories (unadjusted: $p < 0.001$). This association remained after adjusting for covariates (ps < 0.001). Individuals who were poorly monitored were more likely to report ever use of cigarettes by the age of 16 relative to those who were well monitored. These individuals were also more likely to be classed as experimenters and regular smokers (both late-onset and early-onset). Of note, the effect of poor parental monitoring on smoking increased in a relatively linear fashion across the smoking trajectory categories, and was most pronounced in the early-onset regular smoking category (which may be considered the most 'severe' category).

Contrary to expectations, we found no evidence for an association between rs6265 and either smoking initiation phenotype in unadjusted or adjusted analyses (ps > 0.52). No evidence of an association was noted between rs1051730 and ever use of cigarettes ($ps > 0.16$). Curiously, however, there was some weak evidence of an association between this variant and initial smoking trajectories in adjusted analyses (ps > 0.09) , although in the opposite direction to that predicted, with the T allele (associated previously with increased daily cigarette consumption and nicotine dependence) marginally protective for late-onset and early-onset regular smoking.

Table 3.3. Logistic regression models of ever use of cigarettes.

	Unadjusted				Adjusted*			
		95% CI				95% CI		
	OR	LB	UB	*p*	OR	LB	UB	*p*
Parental monitoring	1.53	1.23	1.89	<0.001	1.71	1.36	2.13	<0.001
SNP rs6265 (A)	1.04	0.90	1.19	0.61	1.05	0.91	1.20	0.53
SNP rs1051730 (T)	0.92	0.83	1.03	0.17	0.92	0.82	1.03	0.16

Highest level of parental monitoring used as reference category for parental monitoring. SNPs were coded in terms of number of copies of the minor allele (identified in brackets), '0' being the reference category in both instances. *Results adjusted for sex, housing tenure, overcrowding, smoking during pregnancy, parity, maternal education.

Table 3.4. Logistic regression models of initial smoking trajectories.

	Unadjusted										Adjusted*									
	Experimenters			Late-Onset Regular			Early-Onset Regular			*p*	Experimenters			Late-Onset Regular			Early-Onset Regular			*p*
	OR	*95% CI*		OR	*95% CI*		OR	*95% CI*			OR	*95% CI*		OR	*95% CI*		OR	*95% CI*		
		LB	UB		LB	UB		LB	UB			LB	UB		LB	UB		LB	UB	
Parental Monitoring	1.41	1.06	1.87	1.96	1.46	2.65	2.55	1.67	3.89	<0.001	1.50	1.12	2.01	2.14	1.58	2.92	2.86	1.84	4.46	<0.001
SNP rs6265 (A)	0.97	0.79	1.19	0.91	0.72	1.16	0.99	0.70	1.40	0.89	0.97	0.79	1.19	0.91	0.72	1.15	0.99	0.70	1.40	0.88
SNP rs1051730 (T)	0.98	0.83	1.17	0.84	0.70	1.03	0.79	0.59	1.07	0.19	0.99	0.83	1.17	0.83	0.68	1.02	0.75	0.55	1.02	0.09

Non-smokers used as reference category for smoking trajectory categories. Highest level of parental monitoring used as reference category for parental monitoring. SNPs were coded in terms of number of copies of the minor allele (identified in brackets), '0' being the reference category in both instances. *Results adjusted for sex, housing tenure, overcrowding, smoking during pregnancy, parity, maternal education.

3.3.4 Moderation by parental monitoring

There was no evidence of an interaction between parental monitoring and rs6265 or rs1051730 for either initiation phenotype (*p*s > 0.35).

3.3.5 Sensitivity analyses

Altering the cut-off to distinguish individuals on the basis of high and low parental monitoring level had relatively little impact on our findings. Evidence of an association between parental monitoring and ever use of cigarettes at age 16 years was again observed, although evidence for this was not as strong as that observed for the original cut-off. Those in the low parental monitoring category (comprising 55% of the sample) were more likely to have tried smoking by this age (unadjusted: OR = 1.19, *95% CI* 1.02, 1.38, $p = 0.03$; adjusted: OR = 1.26, *95% CI* 1.08, 1.48, $p = 0.003$). Strong evidence of an effect of parental monitoring on initial smoking trajectories was also observed (unadjusted: $p = 0.006$; adjusted: $p = 0.002$), although again the effects observed were not as strong as those observed for the original cut-off. Those in the low monitoring category (again comprising 55% of the sample) were more likely to be late onset regular smokers (unadjusted: OR = 1.39, *95% CI* 1.07, 1.82; adjusted: OR = 1.46, *95% CI* 1.11, 1.91), and early onset regular smokers (unadjusted: OR = 1.61, *95% CI* 1.07, 2.42; adjusted: OR = 1.68, *95% CI* 1.11, 2.56). Finally, as previously observed, no evidence of an interaction between parental monitoring and rs6265 and rs1051730 was noted for either initiation phenotype (*ps* > 0.43).

3.4 Discussion

We used data from a prospectively assessed cohort to examine evidence for association between rs6265 (*BDNF*) and rs1051730 (*CHRNA3*) and smoking

initiation in adolescence, characterised using two distinct phenotypes: ever use of cigarettes, and initial smoking trajectories. We also sought to determine the potential influence of parental monitoring on these relationships. We found no clear evidence of an association between either genetic variant and either initiation phenotype, nor evidence of modification of genetic effect by parental monitoring in any instance. Strong evidence of an effect of parental monitoring on both initiation phenotypes was observed however.

We found no evidence to suggest that rs6265 in *BDNF* was associated with smoking initiation. This variant was primarily included in this study as a positive control, given its recent association with smoking initiation (Furberg et al., 2010). However, given the sample size of the current study, and the effect size (OR = 1.06) observed for rs6265 in the original meta-analysis, we did not have good power to detect an effect at this locus.

Whilst our small sample size certainly limited our ability to detect very modest genetic effects, we felt that our more precise assessment of the initiation phenotype somewhat offset this issue. Our failure to note such an association in a prospectively assessed cohort calls into question the 'initiation' phenotype typically employed in genome-wide association studies. Issues relating to phenotype definition are discussed below, and in further detail in Chapter 4.

We found strong evidence of an association between parental monitoring and smoking initiation, assessed both in terms of ever use, and initial smoking trajectory. Low parental monitoring was associated with an increased risk of smoking initiation. This finding complements previous research, which has shown that parental monitoring is a risk factor for substance use, including smoking (Bohnert et al., 2009), and emphasises the need to target smoking prevention strategies at this at-risk

group, whilst simultaneously advising parents of the benefits of monitoring their children.

No clear evidence of an association was observed between rs1051730 and smoking initiation. This variant is known to influence receptor response to nicotine (Bierut et al., 2008), which should not plausibly effect very first use of cigarettes. However, one might expect this variant to be associated with behaviours after initial exposure to tobacco, such as risk of progression from experimentation to regular smoking, a pathway captured by our late- and early-onset regular smoker categories. Marginal evidence of an association between this variant was observed for initial smoking trajectories in adjusted analyses ($p > 0.09$) , although in the opposite direction to that predicted, with the T allele (previously associated with increased daily cigarette consumption and nicotine dependence) appearing marginally protective for late-onset and early-onset regular smoking. Reasons for this are unclear. Allele miscoding was ruled out following inspection of allele frequencies. It would certainly be of interest to examine data collected over a longer period into adulthood to capture the development of this influence, ideally combined with a larger sample size.

We found no evidence to suggest an influence of parental monitoring on the relationship between rs1051730 and either initiation phenotype. This finding conflicts somewhat with that of Chen et al. (2009), who observed that the association of rs16969968 with nicotine dependence was modified by level of parental monitoring. Although, unlike Chen et al. (2009), our study was not based on a selected sample, our sample was considerably larger (2687 and 3771 versus 2027 in the Chen et al. study). It is therefore unlikely that our failure to observe a similar

pattern of associations was merely due to low statistical power. One possibility is that the different phenotypes studied here may account for this disparity.

This study has several strengths. The use of data from a prospectively assessed cohort ensured that we could be confident that we were truly measuring smoking initiation. Our use of initial smoking trajectories as a phenotype was also a strength, capturing the complexities of cigarette use during adolescence. This approach may provide richer data for analysis of genetic effects (Munafo & Johnstone, 2008). Characterisation of smoking initiation is normally vague, and certainly varies considerably between studies. Typically, 'ever smokers' are contrasted with 'never smokers' for the purposes of analyses (e.g., Caporaso et al., 2009). Definitions of these two groups vary substantially however. Furberg et al. (2010) contrasted ever *regular* smokers versus never *regular* smokers in their examination of smoking initiation, regular smokers defined as "those who reported having smoked ≥100 cigarettes during their lifetime", never regular smokers defined as "those who reported having smoked between 0 and 99 cigarettes during their lifetime". In contrast, Greenbaum et al. (2006) compared never smokers (reporting never to have smoked a single cigarette in their lives, termed 'noninitiators') with individuals who had "smoked daily for at least 1 year" (termed 'smoking initiators'). Such variation may potentially underlie persistent failures to replicate findings, and highlights the need for more narrowly defined phenotypes, as utilised here. Finally, consideration of important covariates in our analyses, including maternal smoking during pregnancy and level of maternal education, was an additional strength of this study.

Several limitations should be acknowledged. Firstly, our parental monitoring measure relied solely on child self-report, and was assessed using a brief, three item

questionnaire. Ideally we would have also incorporated a parental report, but data on parental monitoring provided by the mothers of the individuals in this sample were not available at (or close to) the age of 11 years. A more comprehensive assessment of parental monitoring developed from Stattin and Kerr's monitoring measure was administered to both ALSPAC offspring and their mothers, the use of which would have enabled us to: a) examine effects of different dimensions of parenting practices on smoking initiation (an issue discussed further below), and b) incorporate both child and parent opinion on monitoring practices. However, these questionnaires were administered when the children were 14 years old, by which time a substantial number had already tried smoking. As such, use of these measures would have required substantial further restriction of our sample, given the necessity to exclude all individuals who had tried smoking before time of assessment. Our results pertaining to level of parental monitoring and smoking initiation are consistent with previous research however, which somewhat negates these concerns. In addition, Laird et al. (2003) have previously observed that robust associations are noted between parental monitoring/knowledge and deviant behaviour regardless of whether either variable is assessed using parental or adolescent report. Secondly, it is important to note that parenting practices encompass a number of different dimensions. In this study we considered only one, namely parental monitoring, primarily focused on parental knowledge of the child's peers and activities with such peers. In contrast, other studies in this field have considered the impact of multiple dimensions. Chen, Storr, & Anthony (2005), for example, examined the impact of parental involvement/reinforcement and coercive parental discipline, alongside parental monitoring, on the risk of exposure to opportunities to try cannabis. Similar patterns of effects were observed for parental monitoring and parental

involvement/reinforcement, namely that low levels of monitoring and involvement were associated with earlier opportunities to try cannabis, and increased risk of said opportunities over the ten year period post assessment. Conversely, low levels of coercive discipline were associated with later opportunities to try cannabis, and decreased risk of said opportunities over the ten year period post assessment, relative to those exposed to higher levels of discipline. This illustrates the importance of considering different aspects of parenting practices. Finally, an additional limitation of our study is that the smoking trajectories characterised did not perhaps capture the full development of smoking behaviours across adolescence. Consideration of smoking data across a broader age range (encompassing adolescence and early adulthood) would allow for a more valid examination of smoking trajectories. Such trajectories are currently being developed within ALSPAC, although were not available for use at the time of writing.

Using data gathered from a prospectively assessed cohort, we examined evidence for association between rs6265 and rs1051730 and two precise measurements of smoking initiation. We also examined the impact of parental monitoring on these relationships. Whilst no evidence of association was observed for either genetic variant, or in support of gene-environment interplay, we do provide evidence to further underscore the importance of parental monitoring in late childhood in predicting risk of smoking initiation in adolescence. Secondly, we also illustrate the potential use of smoking trajectories as a phenotype for use in future examination of genetic effects. A collective move towards the use of such tightly characterised phenotypes may increase likelihood of effect replication. The importance of using more precise, well-defined phenotypes is illustrated and discussed further in the following chapters.

Chapter 4

Laboratory-based Techniques:

CHRNA3, Cotinine and Smoking Topography

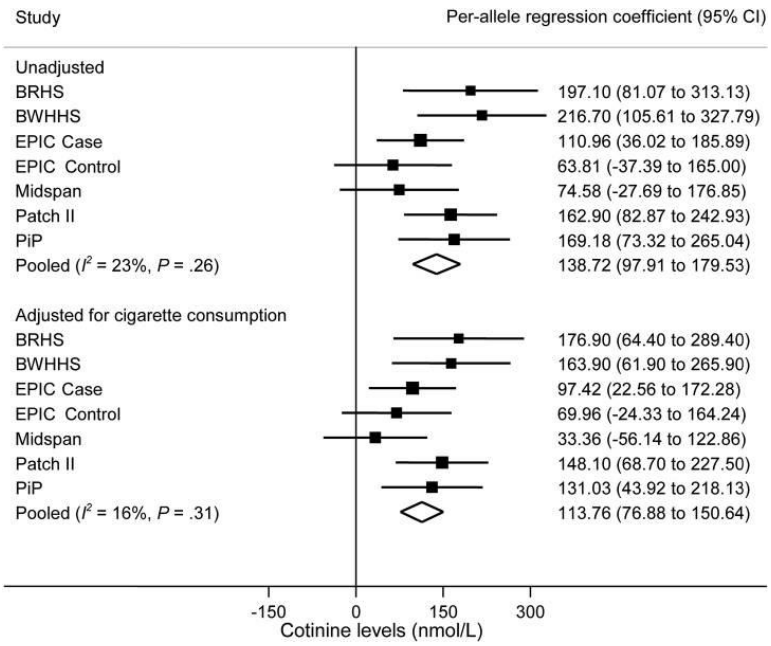
4.1 Introduction

Carefully defined and well-characterised phenotypes offer greater measurement precision, conferring a cleaner genetic signal (through an increase in signal to noise ratio) and improving the likelihood of effect replication. Genome-wide association studies have revealed an association between a locus in the nAChR gene cluster *CHRNA5-A3-B4* and tobacco exposure, crudely defined in terms of self-reported daily cigarette consumption (Furberg et al., 2010; Thorgeirsson et al., 2010). Consequent research has sought to refine this phenotype.

As discussed in Chapter 1, researchers have now begun to examine associations between the 15q locus and objective, precise measures of tobacco exposure, primarily levels of cotinine (the primary metabolite of nicotine) and other nicotine metabolites (Keskitalo et al., 2009; Le Marchand et al., 2008). These preliminary studies have indicated that the risk alleles for heaviness of smoking at this locus are also associated with cotinine levels, and that this association remains after adjustment for self-reported smoking. Munafò et al. (2012) sought to confirm this in a larger sample. Specifically they examined the association between rs1051730 and self-reported cigarette consumption alongside circulating levels of cotinine. Cotinine level was found to show a much stronger association with rs1051730 relative to self-reported cigarette consumption. Moreover, the association between this variant and cotinine was robust to adjustments made for self-reported daily cigarette consumption (see Figure 4.1), reducing the association by only 18%.

This suggests that even among *equal* cigarette consumers, there is genetically influenced variation in total nicotine exposure. Presumably this is due to differences in smoking topography, i.e., how a cigarette is smoked (number of puffs taken per cigarette, volume of smoke inhaled per puff, and so on). It is now well-established that smokers modify their smoking behaviour to self-titrate circulating nicotine to a level appropriate to their need (Strasser et al., 2007). Research using knock-out mouse models suggests that this locus influences self-titrated nicotine exposure via effects at receptors which influence toxicity of high doses nicotine (Fowler et al., 2011).

Figure 4.1. Meta-analysis of association of rs1051730/rs16969968 risk allele with cotinine levels in current smokers (reproduced with permission from Munafo et al., 2012).



Data from six independent studies contributed to the meta-analysis. In each study, linear regression was used to calculate per-allele association of rs1051730-rs16969968 genotype with cotinine levels. Unadjusted and adjusted analyses are shown. Adjusting for daily cigarette consumption had minimal impact on the association.

In collaboration with Nic Timpson (University of Bristol), I developed a protocol to determine potential mediation of the relationship between rs1051730/rs16969968 and cotinine levels by smoking topography. The results of such a study would determine whether the stronger association noted between this variant and cotinine (relative to daily cigarette consumption) is mediated via self-regulated tobacco exposure. A full ethics application for this study was submitted to, and consequently approved by, the ALSPAC Ethics and Law Committee. Unfortunately, due to unavoidable and unexpected delays in participant invitation and circulation, combined with an extremely poor participant response rate of 0.3% (issues which are discussed at length in section 4.5), I was unable to conduct this study before submitting my thesis. Thus, my ability to address the research objective outlined above was prohibited. In the interim period however, I was able to pilot the procedure. Here I present the results of the pilot study, conducted in the absence of genetic data and cotinine data, primarily intended to trial the general procedures and to determine the feasibility of including 'roll-your-own' tobacco smokers in the full study (given concerns regarding the compatibility of this type of cigarette with the smoking topography equipment). An additional, secondary aim was to examine trends in the relationships between self-reported and objectively assessed measures of tobacco exposure. Two methods sections are presented below. The first section presents the protocol for the full study, the second presents details specific to the pilot study.

4.2 Methods: Full study

4.2.1 Experimental design

A recall-by-genotype design will be employed, whereby a genetic variant delivering functional change (in this case rs1051730/rs16969968) is used to select participants for extremely detailed, clinically relevant, phenotype examination. This approach maximises the power and information content of the sample whilst enabling collection of extremely precise phenotypic data impossible to collect in a much larger sample.

4.2.2 Participants

A total of 200 participants will be recruited prospectively from the ALSPAC cohort of mothers on the basis of minor or major homozygote status at rs1051730 (100 in each genotype group). All participants are to be current, daily smokers, in good physical health. Smoking status will be confirmed during initial screening by a carbon monoxide (CO) breath reading (CO breathalyser) and urinary cotinine assessment (yielding a positive or negative reading). Exclusion criteria will include current dependence on any substance other than nicotine and caffeine, and significant current or past physical illness. Pregnant and lactating women are also to be excluded. Participants will be reimbursed for their time with £50 worth of shopping vouchers on completion of the study. Full ethics approval for this study has been granted by the ALSPAC Ethics and Law Committee.

4.2.3 Measures and materials

Smoking topography will be assessed using a smoking topography monitor (CReSS Pocket, Borgwaldt KC, Hamburg) (see Figure 4.2). This is a self-contained, battery-operated device, which measures ambulatory smoking behaviour, with time

and date tags assigned at cigarette insertion/removal, providing a highly quantitative view of cigarette smoking behaviour. Data captured include the following: puff volume; puff duration; puff flow; puffs per cigarette; inter-puff interval; time to first puff; time to removal; volume per cigarette. Onboard memory is used to store all measures. Smoking topography will be assessed both in the laboratory and in the participants' 'natural' environment over the course of one day. Primary outcome measures will be total volume of tobacco smoke consumed per cigarette (ml) and per day (ml).The cigarette smoked in the laboratory is to serve several purposes. Firstly, it will allow participants the opportunity to become familiar with use of the monitor whilst assistance is available. Secondly, it will allow determination of the impact of a quantifiable tobacco dose on cardiovascular and affect measures. Examination of the impact of rs1051730 genotype on these responses may be explored in future research (see Section *5.4.2*).

Cotinine levels are to be assessed from salivary samples. Saliva samples will be collected using salivettes (Sarstedt, Nümbrecht). Samples will be centrifuged twice (at 5800 rpm for 15 minutes) within 24 hours of collection to ensure removal of human tissue, frozen (at -30 ºC) and then sent to ABS Laboratories Ltd. for quantitative analysis of cotinine content.

Genotyping has previously been undertaken by KBioscience Ltd. (www.kbioscience.co.uk), a company who use a proprietary competitive allele specific PCR system (KASPar) for SNP analysis. For this study we will focus on one SNP, namely rs1051730 (*CHRNA3*).

Questionnaires will be used to ascertain demographic information and information on smoking history. The Fagerström Test of Nicotine Dependence (FTND) (Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991) will be used to

determine level of nicotine dependence. The Brief Questionnaire of Smoking Urges (QSU-Brief) (Cox, Tiffany, & Christen, 2001; Tiffany & Drobes, 1991) and the Positive and Negative Affect Scale (PANAS) (Watson, Clark, & Tellegen, 1988) will be administered pre- and post-programmed cigarette smoking to assess craving and affect respectively.

Cardiovascular measures (blood pressure and heart rate) will be assessed using the OMRON M6 blood pressure monitor.

Figure 4.2. Smoking topography monitor (CReSS Pocket, Borgwaldt KC, Hamburg).

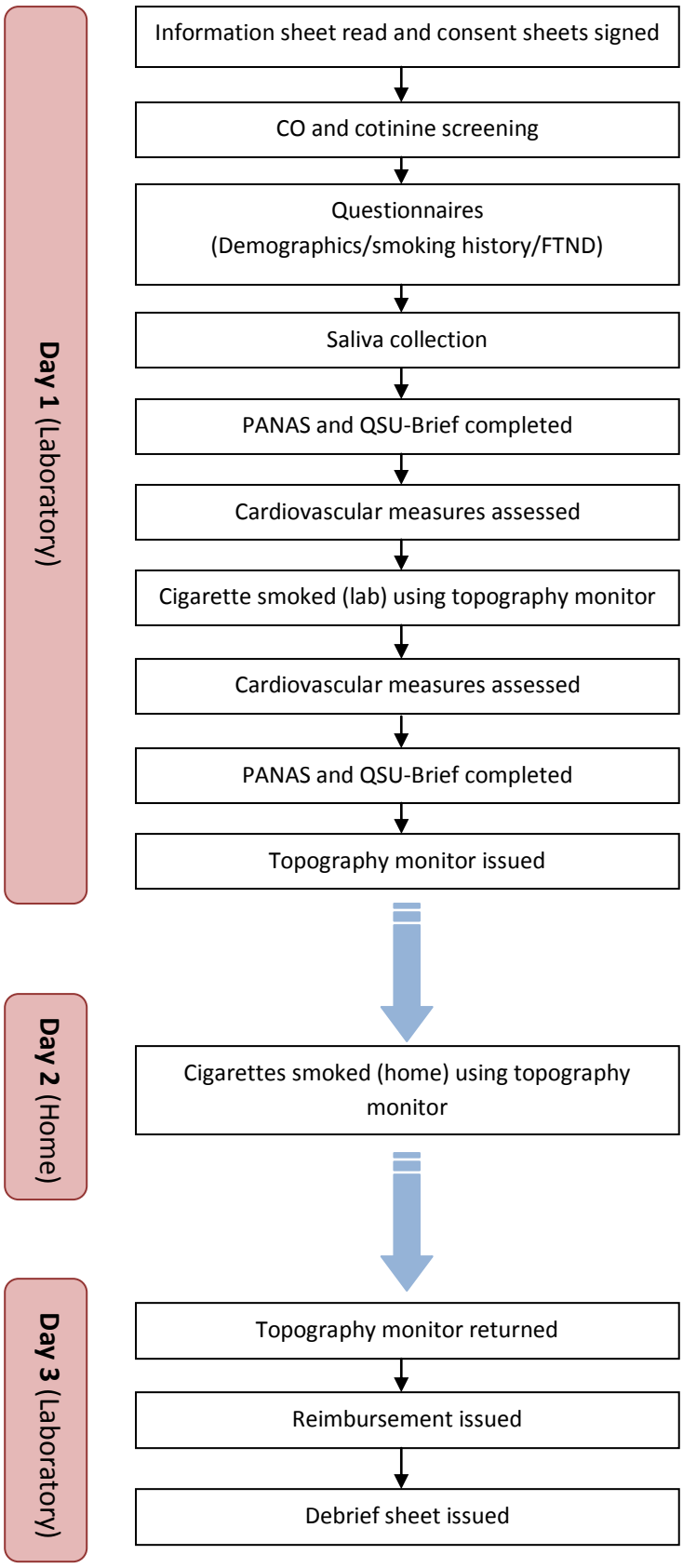
[This image has been removed by the author for copyright purposes]

4.2.4 Procedure

The study will take place over the course of three days (see Figure 4.3). On day one the participant will attend the research centre for approximately 45 minutes. An information sheet will be issued, and written, informed consent provided. Smoking status will be confirmed by a CO breathalyser reading (PiCO+ Smokerlyzer, Bedfont Scientific) and urinary cotinine assessment. A series of questionnaires will then be administered to establish participant demographic information, smoking history, and level of nicotine dependence. A saliva sample will then be collected for quantitative assessment of cotinine level. Participants will consequently be introduced to the smoking topography monitor, and issued with instructions regarding its use. Pre-cigarette PANAS and QSU-Brief questionnaires are to be completed, and cardiovascular measures assessed (blood pressure and heart rate). Participants will then be asked to smoke one of their own cigarettes using the smoking topography monitor. This will take place in a ventilated cubicle (conforming to requirements of the Health Act 2006), under observation by the investigator through one-way glass. The participant will be able to speak to the investigator via an intercom system at any point during the procedure. Post-cigarette PANAS and QSU-Brief questionnaires will then be completed, and cardiovascular measures again assessed. At the end of this session, the participant will be issued with a smoking topography monitor, alongside an information sheet regarding its use and care. On day two, participants will be asked to use the smoking topography monitor for each cigarette consumed that day in their 'natural' environment. On day three, participants will return to the research centre to return the device. Alternatively, if more convenient, the researcher will visit the participant's home

address to collect this. Following completion of this second visit, participants will be reimbursed and a debrief sheet outlining the aims of the study will be issued.

Figure 4.3. Study procedure.



4.2.5 Statistical analysis

Potential mediation of the relationship between rs1051730 genotype and cotinine levels by primary smoking topography outcome measures will be examined. Specifically, cotinine level will be regressed on rs1051730 genotype, to confirm the previously observed relationship. Smoking topography outcome will then be regressed on rs1051730 genotype, to confirm that this variant is indeed a predictor of the mediator. Finally, we will confirm that smoking topography is a significant predictor of cotinine level whilst controlling for rs1051730 genotype.

4.3 Methods: Pilot study

4.3.1 Participants

Participants were recruited from the University of Bristol student body and the general public via circulated advertisements and word of mouth. All were current smokers (confirmed by breath CO reading and/or urinary cotinine assessment), and in good physical and psychiatric health. Exclusion criteria were identical to those outlined above. Participants were reimbursed for their time with £25 upon completion of the study. Full ethics approval for this study was granted by the Faculty of Science Human Research Ethics Committee at the University of Bristol.

4.3.2 Measures and materials

As outlined above. Please note, however, that genetic data were not obtained from these individuals. Please also note that whilst salivary cotinine samples were obtained from individuals participating in the pilot study, quantitative analysis of cotinine content was not conducted (due to laboratory turn-around time combined with imminence of thesis submission deadline).

4.3.3 Procedure

As outlined above. Please note, however, that the second visit took place solely at the research centre for this pilot study.

4.3.4 Statistical analysis

Exploratory analyses were conducted to examine associations between self-reported daily cigarette consumption, FTND score, and smoking topography outcome measures, including actual daily cigarette consumption (Pearson correlation).

4.4 Results

4.4.1 Participant characteristics

Our sample consisted of nine participants, of whom five were male (56%) and four were female (44%). The sample was predominantly White (White: 89%; Asian: 11%). Average age was 26.6 years (range = 20 to 42; *SD* = 7.62). Participants consumed an average of 9.5 cigarettes per day (range = 1 to 20; *SD* = 6.66), as determined through self-report. Just over half of the sample primarily smoked manufactured cigarettes (56%), with a slightly smaller proportion reporting use of 'roll-up' cigarettes (44%) as their primary form of tobacco. A number of participants reported smoking a mixture of both manufactured and 'roll-up' cigarettes. Mean nicotine content of cigarette was 0.66 mg (range = 0.5 to 0.9; SD = 0.22) (NB: these figures could only be calculated for smokers of manufactured cigarettes, as the nicotine content of roll-up cigarettes varies widely due to a number of variable parameters e.g., type of cigarette paper used, diameter of filter, and so on). Mean FTND score was 2.89 (range = 1 to 8; *SD* = 2.32), indicative of relatively low levels of nicotine dependence. Mean age of smoking initiation was 15.4 years (range = 13

to 21; *SD* = 2.95). Mean CO breath reading was 13.11 ppm (range = 3 to 33; *SD* = 9.64).

4.4.2 Smoking topography outcomes

Laboratory: Mean number of puffs taken on the cigarette smoked under observation in the laboratory was 15.3 (95% *CI* 13.5, 17.1; *SD* = 2.3). Mean puff volume was 44 ml (95% *CI* 32, 56; *SD* = 15.2). Over the course of this cigarette, participants inhaled a mean total volume of 671 ml (95% *CI* 482, 861; *SD* = 246.4) (see Table 4.1). For reference, these figures are analogous to those determined in a similar study conducted by Strasser et al. (2007).

Home: Over the course of one day, participants consumed an average of 7.6 cigarettes (95% *CI* 4.5, 10.6; *SD* = 3.9). Mean total number of puffs over the course of one day was 97.4 (95% *CI* 60.8, 134.1; *SD* = 47.7), and mean total volume of smoke inhaled over the course of the day was 5159 ml (95% *CI* 2442, 7876; *SD* = 3534.7) (see Table 4.1).

Per cigarette topography averages were also calculated from data collected at home. On average, participants took 13.1 puffs per cigarette (95% *CI* 12.2, 14.1; *SD* = 1.2), and mean volume of smoke inhaled per cigarette was 668 ml (95% *CI* 482, 855; *SD* = 242.1). These values were comparable to those ascertained in the laboratory (see Table 4.1).

4.4.3 Relationships between measures of tobacco exposure

Scatter plots illustrating relationships between measures of tobacco exposure/nicotine dependence are displayed in Figures 4.4 to 4.7. Evidence of an association between self-reported and objectively assessed daily cigarette consumption was observed ($r = 0.74$, $p = 0.022$; $r^2 = 0.55$). Participants tended to

over-report daily cigarette consumption (although the difference between observed and self-reported mean values was not significant; $p = 0.24$).

Marginal evidence of an association was observed between self-reported daily cigarette consumption and total volume of smoke inhaled over the course of one day ($r = 0.63$; $p = 0.069$, $r^2 = 0.40$). An association was also observed between objectively assessed daily cigarette consumption and total volume inhaled over the course of one day ($r = 0.87$, $p = 0.002$, $r^2 = 0.75$). This second association was unsurprisingly stronger given that these two variables were both recorded on the same day.

Finally, marginal evidence of a correlation was observed between FTND score and total volume of smoke inhaled over the course of one day ($r = 0.67$, $p = 0.051$, $r^2 = 0.45$). Higher FTND scores were associated with larger daily inhalation volumes.

Figure 4.4. Correlation between self-reported and objectively assessed daily cigarette consumption (r = 0.74, *p* = 0.022, r^2 = 0.55).

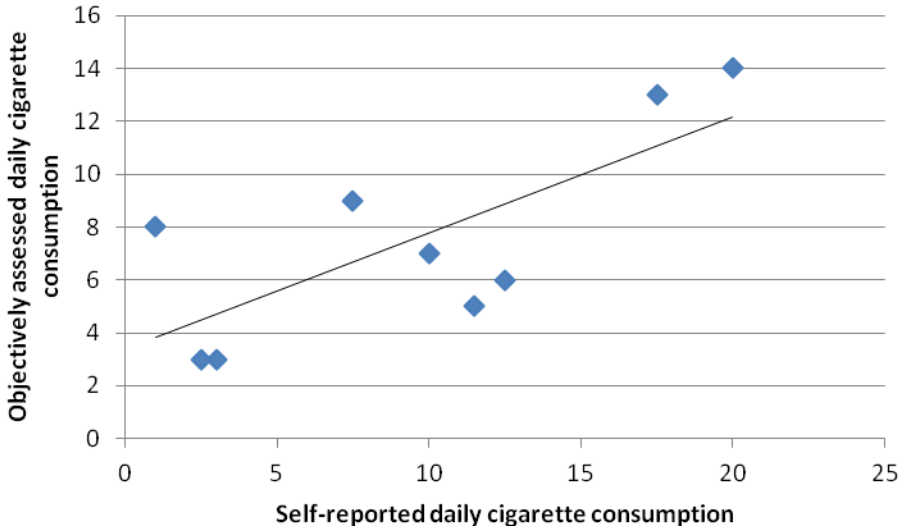


Figure 4.5. Correlation between self-reported daily cigarette consumption and total volume of smoke inhaled over course of day (r = 0.63, *p* = 0.069, r^2 = 0.40).

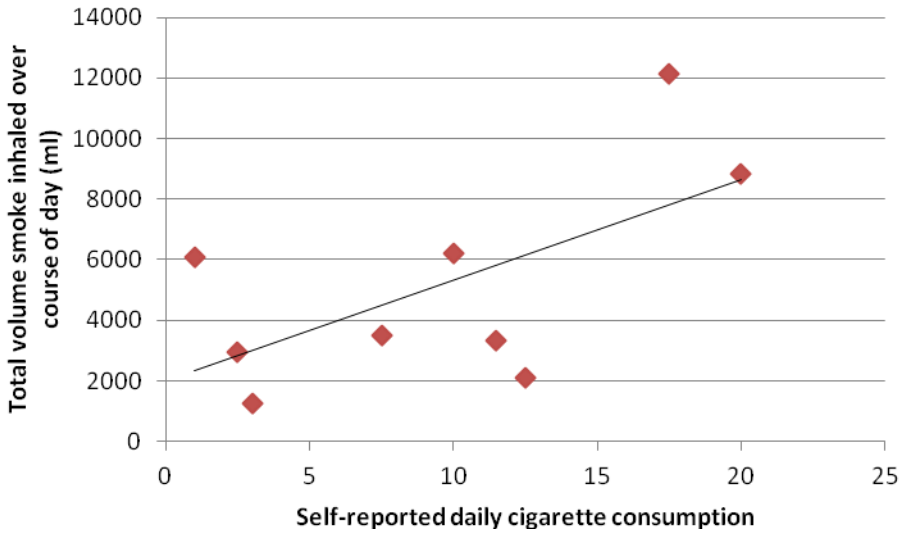


Figure 4.6. Correlation between objectively assessed daily cigarette consumption and total volume of smoke inhaled over course of day (r = 0.87, p = 0.002, r^2 = 0.75).

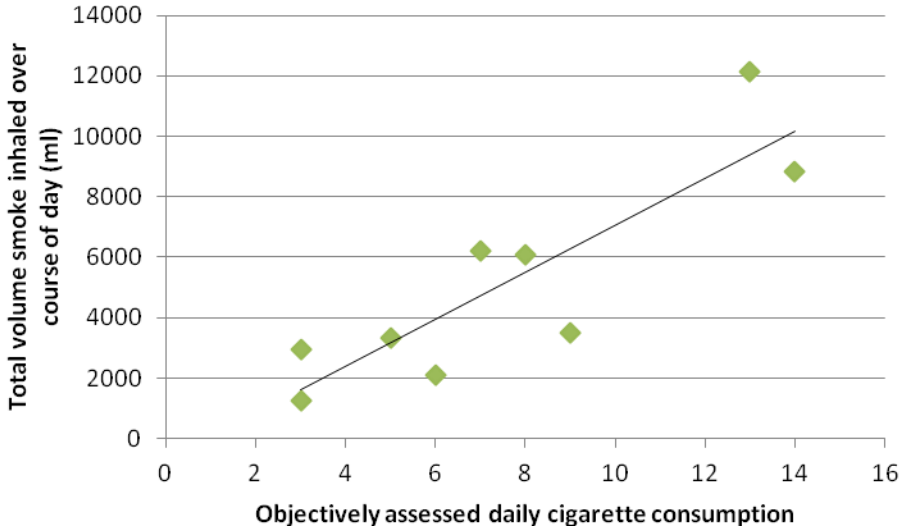
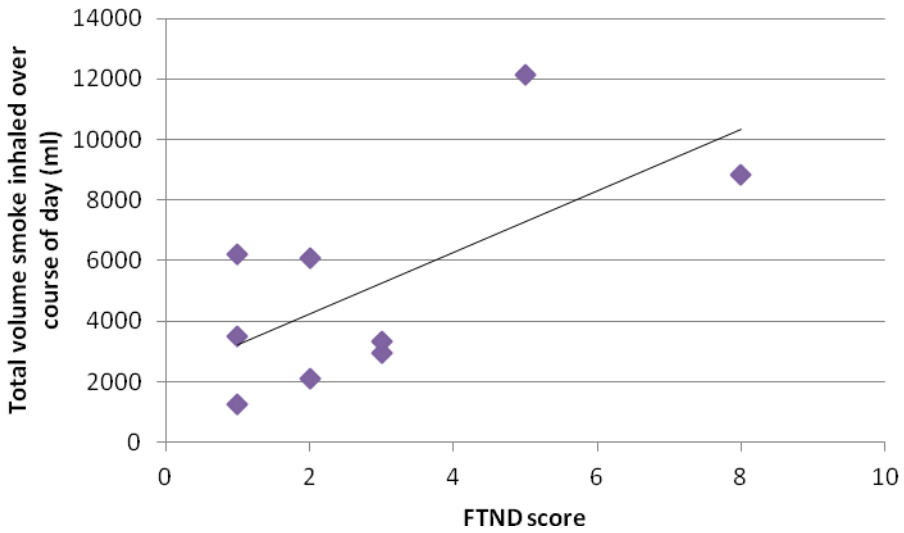


Figure 4.7. Correlation between FTND score and total volume smoke inhaled over course of day (r = 0.67, p = 0.051, r^2 = 0.45).



4.4.4 Impact of participant characteristics on smoking topography outcomes

Smoking topography outcome measures presented according to participant characteristics are displayed in Table 4.1. Analyses of differences in topography outcome measures between participant sub-groups were not conducted given extremely small cell counts which prohibited meaningful statistical assessment.

Table 4.1. Smoking topography outcome measures as a function of participant characteristics.

Participant Characteristics		**Lab topography measures**				**Home topography measures (average per cig)**				**Home topography measures (total over day)**			
	n	Number of puffs		Total volume inhaled (ml)		Mean number of puffs per cig		Mean volume inhaled per cig (ml)		Total number of puffs		Total volume inhaled (ml)	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Overall	9	15.3	2.4	672	246.4	13.1	1.3	669	242.1	97.4	47.7	5159	3534.7
Sex													
Male	5	15.2	3.0	569	210.2	13.4	1.4	646	271.0	106.8	43.5	5458	4021.3
Female	4	15.5	1.7	800	251.7	12.8	1.1	697	237.5	85.8	56.7	4784	3379.1
Cigarette type													
Manufactured	5	15.8	1.1	781	236.5	13.7	0.7	796	228.5	93.4	57.3	5727	4162.8
Roll-your-own	4	14.8	3.5	535	205.8	12.4	1.5	510	161.6	102.5	40.3	4448	3000.3

4.4.5 Cardiovascular and affective responses to cigarette smoking in laboratory

Positive affect decreased following cigarette consumption ($p = 0.009$). Mean change in positive affect (as determined by subtracting post-cigarette positive PANAS score from pre-cigarette positive PANAS score) was -3.4 (95% *CI* -5.8, -1.1*; SD* = 3.0). There was no evidence of a change in negative affect as a consequence of cigarette smoking ($p = 0.43$). Mean change in negative affect (as determined by subtracting post-cigarette negative PANAS score from pre-cigarette negative PANAS score) was -0.6 (95% *CI* -2.1, 0.9, *SD* = 2.0).

Evidence of an increase across all cardiovascular measures was observed as a consequence of cigarette smoking (heart rate: $p = 0.009$; systolic blood pressure: $p = 0.062$; diastolic blood pressure: $p = 0.10$). Mean increase in heart rate was 13.6 bpm (95% *CI* 4.5, 22.7; *SD* = 11.8), mean increase in systolic blood pressure was 7.7 mmHg (95% *CI* -0.5, 15.8*; SD* = 10.6), and mean increase in diastolic blood pressure was 5.0 mmHg (95% *CI* -1.2, 11.2*; SD* = 8.1).

4.5 Discussion

The primary aims of the pilot study were to trial the study procedure and to determine the feasibility of including 'roll-your-own' tobacco smokers in the full study. A secondary aim was to examine relationships between self-reported and objectively assessed measures of tobacco exposure. These objectives were all successfully achieved, and, in relation to our secondary objective, we were able to illustrate the potential importance of using precise, objectively assessed phenotypes when considering tobacco exposure.

One important observation concerned the somewhat weak relationship observed between objectively assessed daily cigarette consumption and total volume of smoke inhaled over the course of one day. Only 75% of the variance noted in total

volume inhaled was accounted for by objectively assessed daily cigarette consumption. Participants who inhaled very similar volumes of smoke over the course of one day varied quite dramatically in daily cigarette consumption. This observation underlines the importance of using precise, objectively assessed measures of tobacco exposure.

Evidence of a relationship between self-reported daily cigarette consumption and objectively assessed daily cigarette consumption was observed. This relationship was not perfect, as predicted. Somewhat unexpectedly however, we found that participants tended to over-report their daily cigarette consumption, rather than under-report. It should be noted however that participants were advised *not* to use the monitor whilst driving, as this would present a safety hazard. As such, the number of cigarettes recorded on the device was likely to under-represent the actual number of cigarettes consumed over the course of a day. Indeed, upon completion of the study a number of participants reported consumption of at least one cigarette outside of the monitor over the course of the home testing day. This was attributed to driving, awkwardness of monitor use, and/or embarrassment of monitor use in public. In two instances the monitor battery died/monitor malfunctioned during the evening of the home testing day, which further contributed to this issue. Given the small effect sizes observed for genetic variants, it is imperative that we obtain information on every cigarette smoked by each participant for the full study. In light of the above, we will ensure that a new battery is inserted into each monitor each week. We will also stress the importance of using the monitor for *every* cigarette that is consumed on the home testing day.

The disparity noted between objectively assessed and self-reported daily cigarette consumption raises an interesting question. Which truly is a more accurate

assessment of heaviness of smoking? Theoretically, the objective assessment should be. However, this is of course dependent on the assumption that people do not change their smoking behaviour whilst using the topography monitor. As we have alluded to above, this does not necessarily hold true. We will be able to provide a more definitive answer to this question in our full study by comparing the relationship between cotinine level (a truly precise, objective assessment of tobacco exposure) and both objectively assessed and self-reported daily cigarette consumption. In either case, we have still illustrated that daily cigarette consumption, however it is assessed, does not account for all the variance noted in total volume of smoke inhaled over the course of a day. Clearly, in addition to the number of cigarettes consumed, the manner in which a cigarette is smoked (number of puffs, puff volume) is crucial here.

A primary aim of this pilot study was to determine the feasibility of including 'roll-your-own' tobacco smokers in our full study. Smoking topography data acquired from roll-up tobacco smokers was somewhat messy. It became apparent from the raw topography data that individuals often had difficulty inserting roll-up cigarettes into the device successfully. This observation was verified by participant feedback upon completion of the study/return of monitor. Such difficulties were rarely observed for manufactured cigarette smokers. It was also apparent that use of the monitor was abandoned by roll-up tobacco smokers on occasion after multiple failed insertion attempts. As such, it appears necessary to exclude roll-your-own tobacco smokers from the full study.

Evidence of changes in cardiovascular and affect measures were observed following cigarette smoking in the laboratory. These figures are presented here solely

for completeness. In the full study, we shall examine the impact of rs1051730 genotype on these responses.

This pilot proved useful for a number of reasons. It enabled us to trial the full study procedure, and also, perhaps most importantly, illustrated the importance of using precise, objectively assessed phenotypes when considering tobacco exposure assessment. This has key implications for epidemiology and genetic association studies, including large genome-wide association studies of smoking behaviour, which typically rely on retrospective self-report measures rather than precise, objective measures of tobacco exposure. As a consequence of this study we will make several amendments to our full study protocol, the most substantial of which entails the exclusion of 'roll-your-own' tobacco smokers. Minor amendments will also be made to case report forms and advice issued to participants concerning home use of the topography monitors.

In addition to piloting the protocol, great effort was also made to conduct the full study. Unfortunately, due to unavoidable delays in participant invitation circulation combined with an extremely poor participant response rate, I was unable to complete this. Following an extensive consultation process with the ALSPAC research team, 360 study invitations were sent to ALSPAC participants in early July 2012. From this initial mail-out, a total of 26 responses were received (7.2%), of which only one was positive. This equates to a positive response rate of 0.3%. These figures stand in stark contrast to those of a similar recall-by-genotype study which was completed recently within the department. This study also entailed the recruitment of participants from the cohort of ALSPAC mothers. Of the 600 invitations sent out for this study, 320 responses were received (53.3%), of which 28 were negative. This equates to a positive response rate of 48.7%. The disparity in

response rate across these two studies may be attributable to a number of factors, perhaps the most major of which concerns the smoking factor. Individuals were not required to smoke in the study presented for comparative purposes, nor was smoking listed in the eligibility criteria. It is possible that individuals do not wish to explicitly hold their smoking habits and behaviours up for examination, particularly in light of the increasing social unacceptability of smoking. An alternative explanation for our poor response rate concerns the time period during which invitations were sent - it is possible that participant availability was poor over the summer months. Finally, it is of note that several ALSPAC sub-studies targeting ALSPAC mothers are currently running, and as such it is possible that this cohort is currently over-burdened.

In light of the difficulties encountered regarding participant recruitment, several amendments will be made to the recruitment process. Future invitation mail-outs will be sent outside of holiday periods. Invitations will also be followed up with postcard reminders (and additional response forms) should no response have been received within two weeks of invite circulation. Furthermore, our target sample may also be widened to include the cohort of ALSPAC fathers (DNA samples are available for ~1000 partners).

Chapter 5

Genome-wide Association Study Meta-analysis:

Association of Multiple Loci with Cotinine Levels in Daily Smokers

5.1 Introduction

The advent of genome-wide technologies has greatly facilitated the search to determine which genetic factors contribute to specific diseases. Genome-wide association (GWA) studies conducted across multiple cohorts/studies (consortia), often featuring total sample sizes in excess of 50,000, are becoming increasingly common, with large sample sizes offering increased power to detect the small genetic effects common in complex behaviours such as smoking. Requiring no *a priori* hypotheses, these studies have proven successful in determining novel variants associated with disease (e.g., Furberg et al., 2010). However, these analyses typically employ relatively crude phenotypes (e.g., self-reported daily cigarette consumption), which is a necessity imposed by the need to harmonise phenotypic definitions across studies.

As previously discussed, objectively assessed phenotypes afford a 'cleaner' genetic signal, and maximise statistical power to detect genetic effects. This has previously been illustrated in relation to tobacco exposure phenotypes, through comparison of the association of rs1051730 (*CHRNA3*) with cotinine levels versus self-reported daily cigarette consumption (Keskitalo et al., 2009; Munafo et al., 2012). Furthermore, the inadequacy of daily cigarette consumption as a proxy for total daily tobacco smoke exposure in current smokers was demonstrated in the previous chapter, focused on smoking topography.

The goal of this chapter was to carry out a multi-centre meta-analysis of GWA data on cotinine levels in current, daily cigarette smokers, in order to identify genetic variants associated with level of tobacco exposure. This study design theoretically maximises statistical power to detect genetic effects, through application of a consortium based approach combined with use of a precise, objectively assessed phenotype.

5.2 Methods

5.2.1 Contributing studies

Seven studies (collectively forming the Cotinine Consortium) contributed to the GWA study meta-analysis: Multi-Ethnic Study of Atherosclerosis (MESA), Coronary Artery Risk Development in Young Adults (CARDIA), Framingham Heart Study, Finn Twin study, Health2000 GenMets study, TwinsUK, and Young Finns Study (YFS) (see Table 5.1). These seven samples resulted in a collective sample size of n = 2,139. The full consortium comprises four additional studies (ALSPAC, Netherlands Twin Registry, FinnRisk 2007, and KORA), although data from these studies were not available for analysis at the time of writing.

5.2.2 Phenotype definition

Cotinine levels were determined from plasma, blood or urine samples, and quantified using immunoassay, radioimmunoassay or mass spectrometry.

5.2.3 Sample inclusion criteria

Individuals within each sample were eligible for inclusion in analyses provided they were assessed for cotinine level at or after 17 years of age, were of European ancestry, were successfully genotyped genome-wide (>95%), and were current *daily* smokers at the time of cotinine assessment. To minimise inclusion of

non-smokers and non-daily smokers in our analyses, specific inclusion thresholds were imposed regarding cotinine level, which we determined on the basis of expert recommendation and ROC analyses. These were conservatively set at 10 ng/ml cotinine in serum/plasma samples assessed using mass spectrometry, 50 ng/ml cotinine in serum/plasma samples assessed using immunoassay, and 250 ng/ml cotinine in urine samples assessed using immunoassay. Full descriptive characteristics of each study participating in the Cotinine Consortium are presented in Table 5.1.

5.2.4 Genotyping and imputation

All contributing studies performed their own genotyping, genotyping quality control, and imputation (see Table 5.2). Studies samples were genotyped on a number of different platforms. Each study applied its own set of quality control filters. Genotype imputation was performed using IMPUTE prior to GWA analyses, using 1000 Genomes (March 2012 release) as a reference, resulting in a common set of approximately nine million SNPs. Unfortunately, imputed data for three samples (CARDIA, MESA and Framingham) was unavailable at the time of writing. As such, only directly genotyped SNPs were included in the meta-analysis for these samples (~500,000).

5.2.5 Study specific GWAS analysis

Prior to study specific GWA analyses, cotinine data were transformed if necessary to correct for positive skew (using natural logarithm or square-root), and then standardised (i.e., converted to Z-scores). An additive genetic model was used for association analyses. Linear regression was used in each instance, with standardised cotinine level as the dependent variable and allele dose (0, 1 or 2 copies

of the minor allele) as the independent variable. Two specific regression models were used:

a) Cotinine = SNP + Sex
b) Cotinine = SNP + Sex + Age

Where age data were not available (i.e., Framingham), analyses using regression model b) were not performed. For family-based studies (e.g., FinnTwin), only one observation per family was included.

5.2.6 Meta-analysis of GWAS results

All GWA study data files were delivered to the University of Bristol via secure file-sharing services. Imputation quality control procedures were then centrally imposed. Specifically, variants were excluded if a) MAF <1%; and/or b) info score < 0.4 or r^2 <0.3. Genomic control correction was applied to all input files prior to running the meta-analysis to correct for population structure. Once the quality of each data file was confirmed, files were imported into METAL (March 2011 Release) (http://www.sph.umich.edu/csg/abecasis/metal/index.html), a software tool for meta-analysis of whole genome association data. A fixed-effects meta-analysis was then performed for each SNP by combining allelic effects weighted by the inverse of their variance. Secondary correction for population structure via genomic control of summary statistics was also performed. The fixed threshold for genome-wide significance was set at $p < 5 \times 10^{-8}$.

5.3 Results

Meta-analysis was completed for 9,736,614 variants. The genomic control parameter (λ_{GC}) for meta-analysis summary statistics was 0.998 so no further adjustments to these data were made. Seventy-one variants exceeded the threshold

for genome-wide significance, set at $p < 5 \times 10^{-8}$ (see Figure 5.1). The ten variants with the lowest *p*-values are displayed in Table 5.3. A QQ plot illustrating the significance of association of all SNPs in the meta-analysis versus that expected under the null hypothesis is displayed in Figure 5.2.

Table 5.1. Descriptive characteristics of Cotinine Consortium studies contributing to the genome-wide meta-analysis.

Study	*n*	Sex (% male)	Age (years)		Cotinine (ng/ml)[1]		Medium	Method
			Mean	SD	Mean	SD		
MESA	207	58.7	59.9	9.1	4235.8	4142.5	Urine	Immunoassay
CARDIA	387	47.8	25.3	3.4	202.0	137.8	Plasma	Radioimmunoassay
Framingham	93	43.0	N/A	N/A	101.3	55.6	Plasma/serum	Mass spectrometry
FinnTwin	145	46.2	23.0	1.5	206.6	107.5	Serum	Mass spectrometry
GenMets	485	57.84	47.3	11.2	490.1	250.6	Serum	Immunoassay
YFS	147	50.3	35.5	3.1	215.7	112.6	Serum	Mass spectrometry
TwinUK*	675	8.9	48.1	13.7	N/A	N/A	Plasma	Unknown

* No unit of measurement available for cotinine currently so unable to impose agreed cut-offs. Smoking status determined on the basis of self-report data in this one instance.

[1] Cotinine mean and SD refer to post-threshold, pre-transformation values.

Table 5.2 Genotyping, imputation and statistical analysis for contributing studies.

Study	Genotyping					Imputation[1]			Association analyses	
	Platform	Inclusion criteria			SNPs met QC criteria	Imputation software	Inclusion criteria		SNPs in meta-analysis	λ_{GC}
		MAF	Call rate	P HWE			MAF	Imputation quality		
MESA	Affy 6.0	.	>99%	.	579,750	N/A	N/A	N/A	504,023[2]	1.006
CARDIA	Affy 6.0	.	>99%	.	720,483	N/A	N/A	N/A	592,178[2]	1.011
Framingham	Affy 5.0	.	>99%	.	500,571	N/A	N/A	N/A	431,492[2]	1.018
FinnTwin*	Illumina 670K	>1%	>95%	>1x10-6	504,770+*	IMPUTE	≥1%	Info > 0.4	8,581,728	1.006
GenMets	Illumina 610K	>1%	>95%	>1x10-6	555,388	IMPUTE	≥1%	Info > 0.4	8,519,598	1.011
YFS	Illumina 670K	≥1%	≥95%	>1x10-6	546,677	IMPUTE	≥1%	Info > 0.4	8,460,465	1.009
TwinUK	Illumina 317K+610K	≥1%	≥95%	>5x10-6	281,269	IMPUTE	≥1%	Info > 0.4	8,392,702	1.012

*This sample comprised three sub-samples, which contributed 549,060, 549,544, and 504,770 directly genotyped variants respectively.
[1]Imputation quality control imposed centrally at University of Bristol.
[2] Directly genotyped SNPs with MAF <1% had not been excluded from these three studies. I excluded these variants before running the meta-analysis, hence values in the 'SNPs in meta-analysis' column differ from those in the 'SNPs met QC criteria' column despite lack of imputation and therefore associated imputation QC procedures.

Table 5.3 Top 10 SNPs associated with cotinine level in genome-wide meta-analysis.

SNP	Chromosome	Gene	Alleles	BETA	SE	*p*	Direction*
rs192004622	3		T/G	-10.539	1.4246	1.38E-13	??????-
rs10519203	15	AGPHD1	T/C	-0.227	0.0323	2.10E-12	-------
rs951266	15	CHRNA5	T/C	0.2364	0.0337	2.20E-12	+?+++++
rs144638935	X		A/G	-7.4124	1.0669	3.72E-12	??????-
rs55853698	15	CHRNA5	T/G	-0.2504	0.038	4.25E-11	???----
rs8034191	15	AGPHD1	T/C	-0.2495	0.0379	4.64E-11	???----
rs2036527	15		A/G	0.2499	0.038	4.90E-11	???++++
rs72740955	15		T/C	0.2497	0.038	4.99E-11	???++++
rs55781567	15	CHRNA5	C/G	-0.2487	0.038	5.89E-11	???----
rs931794	15	AGPHD1	A/G	-0.2486	0.038	5.94E-11	???----

*Refers to direction of effect. A '+' indicates allele 1 is driving the effect, a '-' indicates allele 2 is driving the effect.

Figure 5.1. Manhattan plot illustrating genome-wide association results for the Cotinine Consortium.

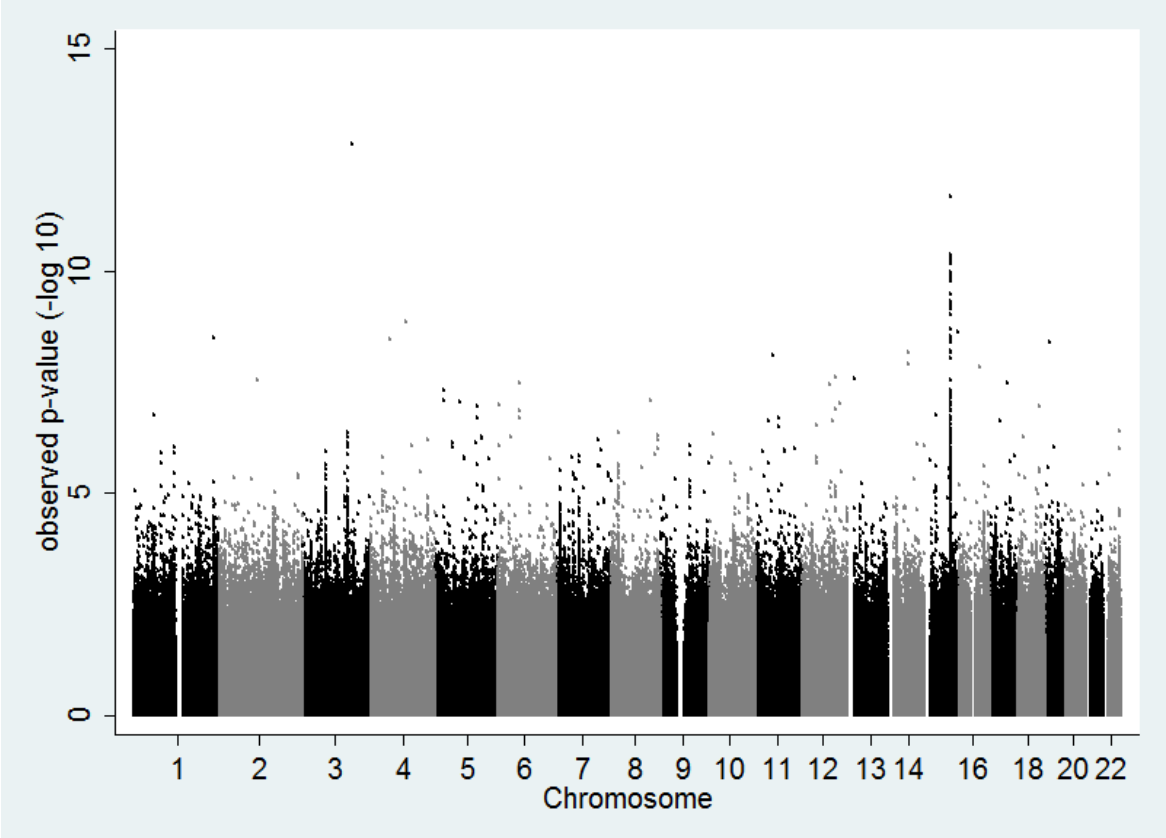
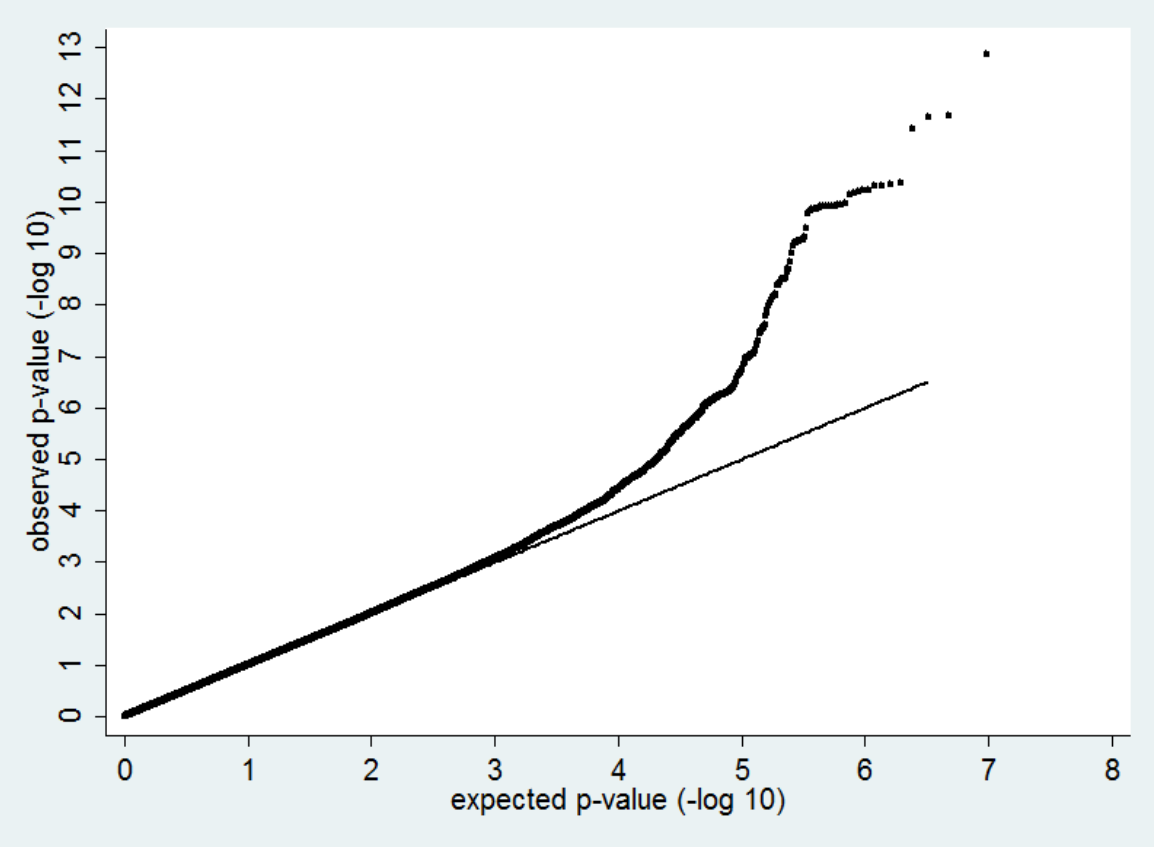


Figure 5.2 QQ plot illustrating significance of association of all SNPs in the genome-wide meta-analysis versus that expected under the null hypothesis.



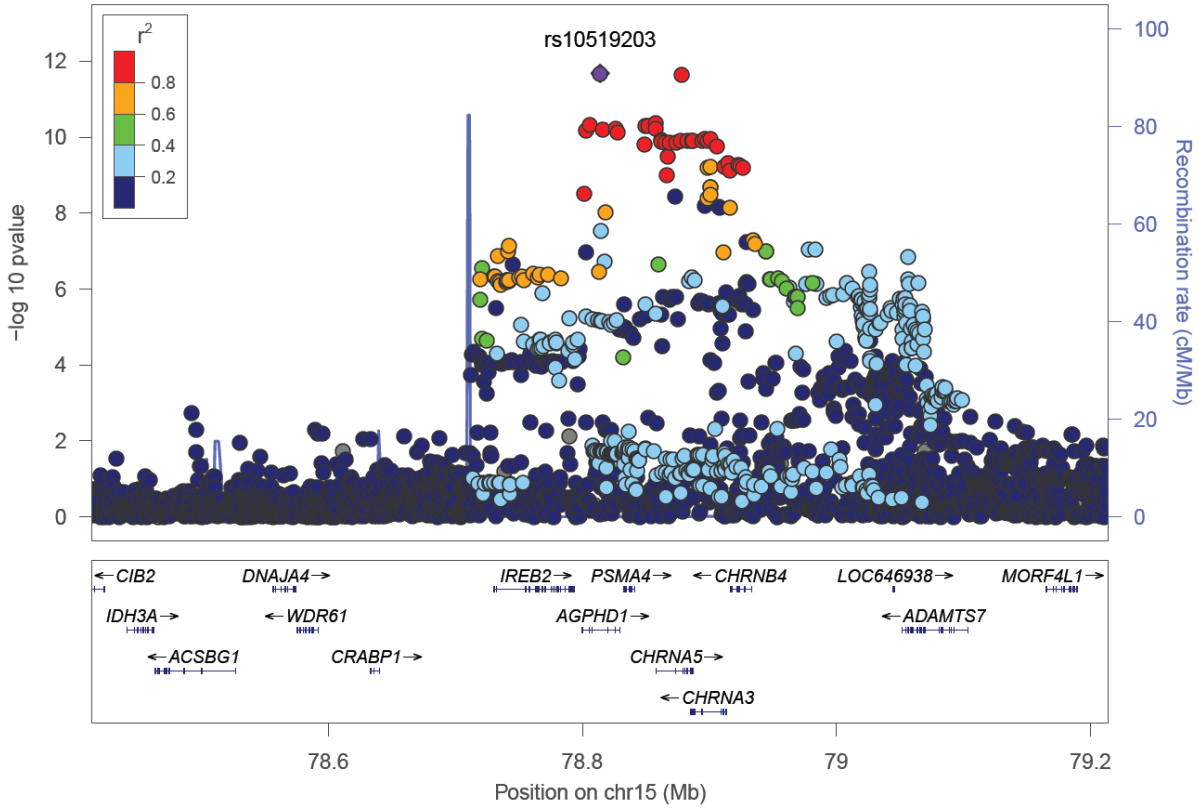
Eight of our top ten SNPs are located within the 15q region on chromosome 15 and are in strong LD with rs1051730, the variant which has formed the focus of this thesis. Our second top hit was rs10519203 with a *p*-value of 2.10×10^{-12}. This variant, located in an intron region of *AGPHD1*, was genotyped and/or imputed successfully in all seven studies (for reference, info score > 0.98 in every instance), and the direction of the effect was consistent across studies (i.e., C allele associated with lower cotinine levels). *AGPHD1*, also known as *LOC123688*, encodes the aminoglycoside phosphotransferase domain containing 1 protein, and is adjacent to the *CHRNA5-A3-B4* gene cluster on chromosome 15 (see Figure 5.3).

The lowest *p*-value generated in our analyses was 1.38×10^{-13} at marker rs192004622. This variant is located in chromosome 3. However, this hit was driven solely by one sample (YFS), having not been genotyped or imputed successfully in the other six samples, and as such should be treated with scepticism. Of note, the info score for this SNP was 0.64, indicative of relatively poor imputation. Similarly, a variant on chromosome X (rs14463893) was also identified as one of our top ten hits. The association noted for this variant was however again driven solely by one study (YFS). The info score for this variant was 0.51, which is again relatively low and suggestive of poor imputation.

Of the remaining 61 variants which reached or exceeded the threshold for genome-wide significance, the vast majority were located in the 15q region of chromosome 15, as illustrated in Figure 5.3. A handful of variants were identified on other chromosomes (namely 1, 2, 3, 4, 6, 11, 13, 14, 16, 17 and 19), although in every instance the association was driven solely by one study, the variant having been unsuccessfully genotyped or imputed in the other six instances. One exception was a variant on chromosome 12, namely rs117376610 ($p = 2.47 \times 10^{-8}$). This SNP

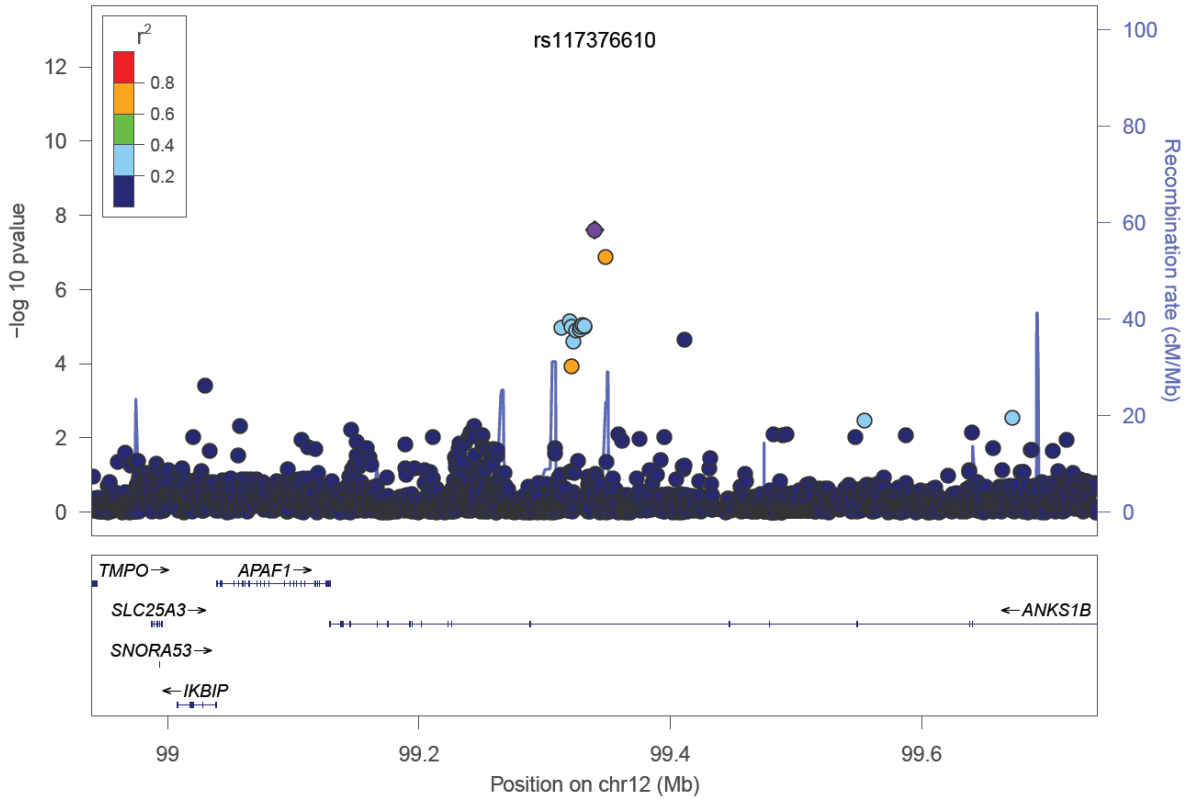
was successfully genotyped and/or imputed in three studies, and the direction of effect was consistent across studies (i.e., A allele associated with higher cotinine levels). This variant is located within *ANKS1B*, a gene which encodes ankyrin repeat and sterile alpha motif domain containing 1B protein. A plot of this region is presented in Figure 5.4.

Figure 5.3. Regional plot of associations with cotinine level on chromosome 15 determined from genome-wide meta-analysis.



SNPs plotted by their position on chromosome 15 against $-\log_{10} p$ value for their association with cotinine level in genome-wide meta-analysis. Estimated recombination rates are plotted in pale blue to reflect local LD structure on secondary *y* axis. The reference SNP (rs10519203) is highlighted in purple. The SNPs surrounding this reference SNP are colour coded to reflect their LD with this variant (see legend). Genome build = hg19; LD population = 1000 Genomes March 2012 release (EUR). Image generated using LocusZoom (http://csg.sph.umich.edu/locuszoom/).

Figure 5.4. Regional plot of associations with cotinine level on chromosome 12 determined from genome-wide meta-analysis.



SNPs plotted by their position on chromosome 12 against $-\log_{10} p$ value for their association with cotinine level in genome-wide meta-analysis. Estimated recombination rates are plotted in pale blue to reflect local LD structure on secondary *y* axis. The reference SNP (rs117376610) is highlighted in purple. The SNPs surrounding this reference SNP are colour coded to reflect their LD with this variant (see legend). Genome build = hg19; LD population = 1000 Genomes March 2012 release (EUR). Image generated using LocusZoom (http://csg.sph.umich.edu/locuszoom/).

5.4 Discussion

We combined a consortium based approach to genome-wide association with use of a precise, objective phenotype, to identify genetic variants associated with level of tobacco exposure. We found strong evidence for an association between a number of variants within the 15q region in chromosome 15 and cotinine level, the majority of which were located in the *CHRNA5-A3-B4* gene cluster and adjacent genes, including *AGPHD1*, and were in strong LD. Additionally, we also found promising evidence for an association with a novel locus in *ANKS1B* on chromosome 12.

The majority of significant associations noted consisted of variants located in chromosome 15, a region robustly linked to a spectrum of tobacco use phenotypes. Our top hit in this region was rs10519203 in *AGPHD1*. This gene, also known as *LOC123688*, encodes the aminoglycoside phosphotransferase domain containing 1 protein, and is adjacent to the *CHRNA5-A3-B4* gene cluster on chromosome 15. The rs10519203 variant is in strong LD with rs1051730 ($r^2 = 0.93$, 1000 Genomes CEU), and has previously been associated with nicotine dependence (Saccone, Wang et al., 2009) and lung cancer risk (Amos et al., 2010). Of note, five of our top ten hits in this region map perfectly onto those identified by Liu et al. (2010) in this region in a genome-wide meta-analysis of smoking quantity. This is encouraging given our sample size of ~2,000 compared to their much larger sample of ~40,000, and illustrates one of the benefits of using precise, objective phenotypes relative to cruder measures. The top hit identified in the 15q region by Liu et al. (2010) corresponded to the fifth top hit in our study, namely rs55853698. This variant is located within the 5' untranslated region of *CHRNA5*, and as such is a potential candidate for affecting mRNA transcription, as Liu et al. (2010) have previously reported.

One promising, unexpected hit observed in this study was variant rs117376610, located in the intron region of *ANKS1B*. *ANKS1B* encodes a multi-domain protein that is predominantly expressed in brain and testis. Interestingly, this gene has previously been found to associate with addiction phenotypes, including polysubstance use, alcohol dependence, and methamphetamine use (Liu et al., 2006; Uhl et al., 2008). Also of potential interest is that down-regulation of this gene has been noted in oral cancer tissues (Zain et al., 2010). However, to the best of our knowledge, this gene has not previously been associated with heaviness of smoking phenotypes.

It is of note that this chapter serves as an interim presentation of our results. To date, 11 studies (comprising 12 samples) have agreed to contribute to the Cotinine Consortium. Whilst GWA data have only been provided from seven studies thus far, the results of which are included here, we shortly expect to receive additional data from our remaining samples. Once data from all constituent studies have been received, alongside data relating to additional, imputed variants from the CARDIA, MESA and Framingham studies, we will re-run our meta-analysis. The increased combined sample size should theoretically afford additional power to detect further variants. It will also be of certain interest to see if the significant association in *ANKS1B* holds in this larger sample. Furthermore, we also plan to replicate our findings in an independent sample if possible.

In summary, using a genome-wide meta-analytic approach, we have found evidence for association between cotinine levels in current smokers and, on the one hand, the 15q region on chromosome 15 (a locus which has been found to robustly associate with tobacco exposure in previous studies), and on the other, a potential novel locus located in *ANKS1B* on chromosome 12. In addition, we have illustrated

how the use of precise, objective phenotypes in GWA studies allows for the replication of findings generated by studies with much larger sample sizes. We hope to identify additional, novel variants as the total sample size of the consortium expands.

Chapter 6

Discussion

6.1 Summary of main findings

6.1.1 Systematic review and meta-analysis (Chapter 2)

The aim of this study was to evaluate the strength of evidence for the association between rs1051730 (*CHRNA3*) and rs16969968 (*CHRNA5*) and heaviness of smoking, assessed in terms of daily cigarette consumption. Secondary aims were to determine which (if either) of the two variants provided a stronger genetic signal, test for the existence of small study bias, explore the impact of year of publication, and investigate the impact of ancestry and disease state as potential moderating variables. Meta-analysis indicated compelling evidence of an association between the rs1051730/rs16966968 variants and daily cigarette consumption, equivalent to a per allele effect of approximately one cigarette per day. Weak evidence of small study bias was observed, although adjustment for this had minimal effect on the overall effect estimate. No evidence of an association between effect size estimate and year of publication was observed. The genetic variant rs1051730 was found to provide a stronger signal than rs16966968 in stratified analyses, although this difference was only qualitatively observed in the subset of samples which provided data on both variants. No other differences in effect sizes were observed between stratified groups. In short, the rs1051730/rs16969968 locus is unequivocally associated with smoking, although uncertainty remains with respect to both the mechanism underlying this association, and other phenotypes with which this variant might be associated.

6.1.2 Genetic epidemiology (Chapter 3)

The aim of this study was to examine the association between rs1051730 and smoking initiation in adolescence, characterised using two distinct phenotypes: 'ever use' of cigarettes, and initial smoking trajectories, the latter determined using repeated measures of smoking frequency. This was examined using data from a prospectively assessed cohort. The potential influence of parental monitoring on this relationship was also examined. No evidence for association between rs1051730 and either initiation phenotype was observed, nor any evidence for effect modification by parental monitoring. The predicted main effect of parental monitoring on smoking initiation was observed. In conclusion, rs1051730 does not appear to be strongly associated with smoking initiation, although the relatively small sample size limited ability to detect modest genetic effects. While parental monitoring is strongly associated with smoking initiation, it does not appear to modify any association between rs1051730 and initiation, although this conclusion is limited by the failure to observe a main effect of this genotype.

6.1.3 Laboratory-based techniques (Chapter 4)

The objective of this section was to examine potential mediation of the relationship between rs1051730/rs16969968 and cotinine levels by smoking topography. The results of such a study would determine whether the stronger association noted between this variant and cotinine (relative to daily cigarette consumption) is mediated via self-regulated tobacco exposure. Due to unavoidable and unexpected delays in participant invitation, combined with an extremely poor response rate, I was unable to conduct this study. A pilot study was however completed, conducted in the absence of genetic data and cotinine data, primarily intended to trial the protocol procedures, and to examine associations between both

objectively assessed and self-reported measures of tobacco exposure. The viability of including 'roll-your-own' tobacco smokers in the full study was also assessed. This pilot served as a useful feasibility study. It also, perhaps most importantly, illustrated the importance of using precise, objectively assessed phenotypes when considering tobacco exposure assessment. This has key implications for epidemiology and genetic association studies, including large genome-wide association studies of smoking behaviour, which typically rely on retrospective self-report measures rather than precise, objective measures of tobacco exposure.

6.1.4 Genome-wide meta-analysis (Chapter 5)

The primary objective of this project was to identify genetic variants robustly associated with heaviness of smoking, in this instance assessed in terms of cotinine level, which, as has previously been demonstrated, is a more precise and objective assessment of tobacco exposure relative to self-reported daily cigarette consumption, thus offering a 'cleaner' genetic signal, and maximising power to detect genetic effects. To this end, we created a consortium ('Cotinine Consortium') comprised of seven studies to conduct a genome-wide meta-analysis of this phenotype. As expected, we found strong evidence for an association between a number of variants within the 15q region in chromosome 15 and cotinine level, the majority of which were located in the *CHRNA5-A3-B4* gene cluster and adjacent genes, including *AGPHD1*, and were in strong LD. Additionally, we also found promising evidence for an association with a novel locus in *ANKS1B* on chromosome 12. From a methodological perspective, we also illustrated how the use of precise, objective phenotypes in GWA studies allows for the replication of findings generated by studies with much larger sample sizes, and also potential allows for the identification of additional, novel variants.

6.2 Research implications

The robust association we, and others, have demonstrated between rs1051730 and heaviness of smoking has definite practical implications for future research. Given said association, rs1051730 is well-suited to serve as an instrumental variable in Mendelian randomisation studies (see Davey Smith & Ebrahim, 2003), in this instance allowing the investigation of causal links between tobacco smoking and a variety of outcomes. Its utility in this context has previously been demonstrated (e.g., Freathy et al., 2011), and will be extended upon during the fellowship I have recently been awarded (Oak Foundation), in which I will be investigating causal links between tobacco use and a variety of health outcomes (including anxiety and depression).

The pilot smoking topography study demonstrated the superiority of precise, objectively assessed tobacco exposure phenotypes (i.e., inhalation volumes) over the more traditionally utilised phenotypes such as daily cigarette consumption (self-reported *or* objectively assessed). Perhaps the most pertinent illustration of this was the observation that individuals who inhaled the same volume of smoke over the course of one day ranged widely in terms of daily cigarette consumption. This has key implications for epidemiology and genetic association studies, including large genome-wide association studies of smoking behaviour, which typically rely on retrospective self-report measures rather than more precise, objective measures of tobacco exposure. Our consequent adoption and application of this approach to the GWA field further illustrated the benefits of using precise, objective phenotypes relative to cruder measures.

On a related note, the genetic epidemiological study presented in Chapter 3 illustrated the potential use of smoking trajectories as a phenotype for use in future

genetic association studies. A collective move towards the use of such tightly characterised phenotypes may increase the likelihood of effect replication. This statement should however be considered alongside the counter argument that simple phenotypes (with fewer assessment points and no requirement for data imputation and complex statistics) are easier to replicate than those that are complex. Simultaneously, if a genetic variant influences a 'top-level' phenotype (e.g., ever smoking) through multiple pathways, then the dissection and analysis of sub-level phenotypes is unlikely to lead to effect replication (given even smaller genetic effect sizes).

6.3 Limitations

Minor limitations relating to each of the studies included in this thesis have been acknowledged and discussed in the preceding experimental chapters. This section focuses on the major limitations of each of these studies, alongside much broader limitations inherent to these fields of study, which impact both the studies included in this thesis and the wider literature.

Meta-analysis (Chapter 2): The primary limitation of the meta-analysis was that I did not have the data necessary to perform a joint SNP analysis of rs1051730 and rs16969968, in which the effects of one variant were conditioned on the other. This analysis would have enabled me to comment more authoritatively on the difference in genetic signal between these two variants, if any, which are known to be in linkage disequilibrium. In addition, the focus of this study centred on what is essentially one genetic signal, whereas we now know that there are several independent signals within this region, all contributing to a proportion of phenotypic variance (for example, see Liu et al., 2010).

Genetic epidemiology (Chapter 3): No evidence of an association was observed between rs6265 in *BDNF* and smoking initiation. This variant was included as a positive control, and its failure does somewhat limit the conclusions that can be drawn regarding observations for rs1051730. The sample upon which this genetic epidemiological study was based was relatively small, thus limiting ability to detect very modest effects, such as that identified between rs6265 and smoking initiation by Furberg et al. (2010). Unfortunately there was little we could do to avoid this given sample attrition over time (an issue common to longitudinal studies) and variable requirements. A potential solution to this problem would involve combining ALSPAC data with comparable data from an additional birth cohort study, a larger sample offering increased power to detect such effects (see section *6.4.1*). That said, we felt that our assessment of the initiation phenotype, which was much more precise than that typically employed in large GWA studies, somewhat offset this issue.

Laboratory-based techniques (Chapter 4): Unavoidable delays in participant invitation circulation combined with an extremely poor participant response rate precluded completion of the full planned recall-by-genotype smoking behaviour study. Unfortunately this prohibited examination of, or ability to comment on, potential mediation of the relationship between rs1051730 genotype and cotinine level by smoking topography measures. However, a full trial of the protocol was conducted, which illustrated an important point concerning precision of smoking-related phenotypes. We were unable to comment conclusively on the superiority of objectively assessed daily cigarette consumption over consumption determined through self-report however in the absence of cotinine data. This will be examined in the full study which will be conducted during my fellowship.

Genome-wide meta-analysis (Chapter 5): There were no major limitations to report in relation to this study. Whilst the combined sample size employed was relatively small for a genome-wide meta-analytic approach, our findings were almost wholly analogous to those identified in studies twenty times the size. This boldly illustrates one of the benefits of using precise, objective phenotypes relative to cruder measures.

6.4 Future directions

6.4.1 Genetic association studies

Progress in the identification of genetic variants *robustly* associated with smoking-related phenotypes has been limited (the *CHRNA5-A3-B4* locus is a rare exception). This is perhaps unsurprising given that the effects attributable to specific genetic variants tend to be very small in complex behaviours such as smoking. Use of crudely defined phenotypes, typically based on self-report, further confounds this issue. Future studies should seek to use much larger, consortia-based samples, maximising statistical power to detect effects, ideally combined with precisely defined, objective phenotypes, which offer a 'cleaner' genetic signal, an approach exemplified in Chapter 5. In addition, replication of studies should be positively encouraged. These studies should be adequately powered, rather than simply reflecting the size of the original 'discovery' study sample. Submission and publication of null findings (from adequately powered studies) should be encouraged, in relation to both novel studies, and replications.

6.4.2 Recall-by-genotype studies

Recall-by-genotype studies involve the selection of participants on the basis of genotype at a specific locus delivering functional change. This approach

maximises the power and information content of the sample whilst enabling collection of extremely precise phenotypic data impossible to collect in a much larger sample, which may assist in the identification of potential mechanisms underlying genetic associations. The first logical extension of this thesis involves completion of the planned recall-by-genotype smoking topography study, examining potential mediation of the relationship between rs1051730 genotype and cotinine level by smoking topography measures.

In a similar vein, it would also be of interest to examine the association between this locus and objectively assessed responses to nicotine challenge, again utilising a recall-by-genotype based approach. As discussed, research using knock-out mouse models suggests that this locus influences self-titrated nicotine exposure via effects at receptors which influence toxicity of high doses of nicotine, particularly those localised to the medial habenula - interpeduncular nucleus pathway (Fowler et al., 2011). Translational research seeking to evaluate the effect of nicotine challenge on brain activation as a function of rs16969968 genotype using neuroimaging technologies is now called for, which may point to new targets for novel smoking cessation therapies. Additional laboratory based techniques/measures which could be used to investigate this aim include cardiovascular measures, galvanic skin response, and questionnaire batteries, administered pre- and post-administration of nicotine.

6.4.3 Phenotype refinement

The methods developed and employed within this thesis, focused on careful refinement of phenotype, hold the potential for wider application. They could, for example, be extended to examine cannabis use. Loci robustly associated with cannabis use and dependence have yet to be identified. It is likely, however, that the effects we will observe between genetic variants and cannabis use will operate in a

similar manner to those observed for tobacco. If this is the case, then precise, objective measures of cannabis use (e.g., circulating metabolite levels of tetrahydrocannabinol and/or cannabindiol) may improve the success of GWA studies in identifying variants associated with cannabis use. Firstly, however, we would need to establish whether or not cannabis smoking is plastic in a similar way to tobacco smoking (as illustrated by Strasser et al., 2007). Specifically, we would need to examine whether cannabis smokers modify their smoking behaviour to self-titrate psychoactive components of cannabis to a level appropriate to their need. Precise dissection of this phenotype would be of certain benefit to large GWA studies seeking to determine variants associated with cannabis use.

6.4.4 Clinical applications

From a clinical perspective, the ultimate goal of this line of research is to improve the efficacy of smoking cessation treatment. This may be accomplished via two avenues: 1) identification of novel treatment targets; and 2) genetic tailoring of existing pharmacotherapies ('personalised medicine'). In this section we discuss how our research, alongside other relevant advancements in the field, has contributed towards these goals.

Novel treatment targets: A robust association between the genetic variant rs1051730/rs16969968 and heaviness of smoking is now firmly established, as we and a number of other groups have demonstrated (e.g., Furberg et al., 2010; Ware, van den Bree, & Munafo, 2011). Perhaps counter to the usual route of scientific inquiry, these exciting findings, based exclusively on human samples and strengthened by their identification through agnostic genome-wide methods, have led to preclinical research focused on determining the mechanism underlying these associations. Exciting progress has been made using knockout mouse models,

highlighting the importance of α5 nAChR subunits in regulating nicotine intake, particularly those localised to the MHb-IPN pathway. Translational research seeking to evaluate the effect of nicotine challenge on brain activation as a function of rs16969968 genotype using neuroimaging technologies is now called for, which may point to new targets for novel smoking cessation therapies. It is possible, for example, that pharmacological stimulation of the MHb-IPN tract may act to reduce/limit nicotine intake/tobacco use.

Personalised medicine: Given the robust association noted between the 15q locus and heaviness of smoking, one might speculate that individuals carrying the rs1051730/rs16969968 risk variant would benefit from an increased dose of nicotine replacement therapy (NRT) relative to non-carriers (to assist quit attempts). It is important to note, however, that the locus in question only accounts for a very small proportion of variance in this phenotype, and, over and beyond this, it may well be as effective to universally increase NRT dose, regardless of genotype. This latter option may certainly prove more economical. There is, however, a substantial genetic component to nicotine dependence (~.50). Therefore, the development of genetic risk scores based on multiple genetic variants (accounting for a much larger proportion of variance in heaviness of smoking) may hold more promise with regards to implementing such an approach.

To date, the major pharmacogenetic success stories have come about for treatments with severe and unpleasant side effect profiles (e.g., chemotherapy treatments). As such, a pharmacogenetic approach to smoking cessation is perhaps most likely to be productive in identifying people who can tolerate second-line treatments (e.g., clonadine), which, whilst effective, aren't widely used due to high

incidence of adverse effects such as sedation and nausea (Gourlay, Stead, & Benowitz, 2004).

As an aside, it is important to consider whether or not genetic tailoring of medication may prove to be cost-effective in relation to universal treatment in the field of smoking cessation. This has certainly been called into question (Ware & Munafo, 2012). Moreover, should this approach eventually prove more economical and effective than 'broad-stroke' medicine, we are still faced with important ethical issues and questions regarding the practicality of clinical implementation. The potential for discrimination (by employers and health insurers) on the basis of genetic information is certainly a key ethical concern (Shields, Lerman, & Sullivan, 2004). From a more practical standpoint, it is also important to consider the ability of health care providers to relay such information. Are general practitioners and other front-line clinicians sufficiently knowledgeable and adequately trained to provide information to patients regarding genetically tailored treatment options? And are they able to provide genetic counselling? An appropriate level of understanding (alongside availability of time and resources) will be required of clinicians delivering such information. These issues are discussed at length in Shields et al. (2004). The genetic literacy of the general public is also to be considered, and there is certainly a call for quantitative and qualitative research examining patient response to genetically tailored treatment. This is an area which is beginning to make progress (BEACON trial, Stanford University; http://clinicaltrials.gov/ct2/show/NCT00991081). We must also question the potential for the application of pharmacogenetic treatment approaches in the resource-limited developing world, where prevalence of tobacco use is increasing (WHO, 2008).

6.5 Conclusions

The primary aim of this thesis was to determine the role played by the *CHRNA5-A3-B4* gene cluster in smoking-related behaviours, with an emphasis on phenotype refinement to aid understanding of the mechanisms underlying these associations. A number of different approaches were utilised to address this objective, namely systematic review and meta-analysis, genetic epidemiology (including detailed phenotyping of smoking behaviour in adolescence), laboratory-based techniques, and genome-wide meta-analysis.

We found compelling evidence for a small, robust association between the rs1051730/rs16966968 variants and daily cigarette consumption, equivalent to a per allele effect of approximately one cigarette per day. This effect was consistent across population sub-groups. Compelling evidence for an association between this locus and level of tobacco exposure was further illustrated through genome-wide meta-analysis of cotinine levels in current smokers. No association was observed between this locus and smoking initiation however, as examined in a prospectively assessed cohort using precisely defined phenotypes, although this observation should be viewed tentatively in light of the failure of our positive control. An association between rs1051730/rs16969968 and smoking topography has yet to be reported. However, a full protocol was developed and piloted to investigate this. In addition, we have also illustrated the importance of precise, objective, phenotype definition, an observation which has important implications for the fields of molecular genetics and epidemiology.

References

Amos, C. I., Gorlov, I. P., Dong, Q., Wu, X., Zhang, H., Lu, E. Y., et al. (2010). Nicotinic acetylcholine receptor region on chromosome 15q25 and lung cancer risk among African Americans: a case-control study. *Journal of the National Cancer Institute, 102*(15), 1199-1205. doi:10.1093/jnci/djq232

Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T., et al. (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genetics, 40*(5), 616-622. doi:10.1038/ng.109

Baker, T. B., Weiss, R. B., Bolt, D., von Niederhausern, A., Fiore, M. C., Dunn, D. M., et al. (2009). Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine and Tobacco Research, 11*(7), 785-796. doi:10.1093/ntr/ntp064

Balfour, D. J. (2002). The neurobiology of tobacco dependence: a commentary. *Respiration, 69*(1), 7-11. 49362 [pii]

Benowitz, N. L. (2008). Neurobiology of nicotine addiction: implications for smoking cessation treatment. *American Journal of Medicine, 121*(4 Suppl 1), S3-10. S0002-9343(08)00103-4 [pii]
10.1016/j.amjmed.2008.01.015

Benowitz, N. L., Herrera, B., & Jacob, P., 3rd. (2004). Mentholated cigarette smoking inhibits nicotine metabolism. *Journal of Pharmacology and Experimental Therapeutics, 310*(3), 1208-1215. 10.1124/jpet.104.066902
jpet.104.066902 [pii]

Bierut, L. J. (2009). Nicotine dependence and genetic variation in the nicotinic receptors. *Drug and Alcohol Dependence, 104*(SUPPL. 1), S64-S69.

Bierut, L. J., Stitzel, J. A., Wang, J. C., Hinrichs, A. L., Grucza, R. A., Xuei, X., et al. (2008). Variants in nicotinic receptors and risk for nicotine dependence. *American Journal of Psychiatry, 165*(9), 1163-1171. doi:10.1176/appi.ajp.2008.07111711

Bohnert, K. M., Rios-Bedoya, C. F., & Breslau, N. (2009). Parental monitoring at age 11 and smoking initiation up to age 17 among Blacks and Whites: a prospective investigation. *Nicotine and Tobacco Research, 11*(12), 1474-1478. doi:10.1093/ntr/ntp160

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., et al. (2012). Cohort Profile: The 'Children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*.

Breetvelt, E. J., Numans, M. E., Aukes, M. F., Hoeben, W., Strengman, E., Luykx, J. J., et al. (2011). The association of the alpha-5 subunit of the nicotinic acetylcholine receptor gene and the brain-derived neurotrophic factor gene with different aspects of smoking behavior. *Psychiatric Genetics*. doi:10.1097/YPG.0b013e32834c0c75

Breitling, L. P., Dahmen, N., Mittelstraß, K., Illig, T., Rujescu, D., Raum, E., et al. (2009). Smoking cessation and variations in nicotinic acetylcholine receptor subunits α-5, α-3, and β-4 genes. *Biological Psychiatry, 65*(8), 691-695. doi:10.1016/j.biopsych.2008.10.004

Broderick, P., Wang, Y., Vijayakrishnan, J., Matakidou, A., Spitz, M. R., Eisen, T., et al. (2009). Deciphering the impact of common genetic variation on lung

cancer risk: A genome-wide association study. *Cancer Research, 69*(16), 6633-6641. doi:10.1158/0008-5472.CAN-09-0680

Caporaso, N., Gu, F., Chatterjee, N., Sheng-Chih, J., Yu, K., Yeager, M., et al. (2009). Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One, 4*(2), e4653. doi:10.1371/journal.pone.0004653

Carter, L. P., Stitzer, M. L., Henningfield, J. E., O'Connor, R. J., Cummings, K. M., & Hatsukami, D. K. (2009). Abuse liability assessment of tobacco products including potential reduced exposure products. *Cancer Epidemiology Biomarkers and Prevention, 18*(12), 3241-3262. 18/12/3241 [pii]
10.1158/1055-9965.EPI-09-0948

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science, 301*(5631), 386-389. 10.1126/science.1083968
301/5631/386 [pii]

Changeux, J. P. (2010). Nicotine addiction and nicotinic receptors: lessons from genetically modified mice. *Nat Rev Neurosci, 11*(6), 389-401. nrn2849 [pii]
10.1038/nrn2849

Chen, C. Y., Storr, C. L., & Anthony, J. C. (2005). Influences of parenting practices on the risk of having a chance to try cannabis. *Pediatrics, 115*(6), 1631-1639. 115/6/1631 [pii]
10.1542/peds.2004-1926

Chen, D., Truong, T., Gaborieau, V., Byrnes, G., Chabrier, A., Chuang, S. C., et al. (2011). A sex-specific association between a 15q25 variant and upper aerodigestive tract cancers. *Cancer Epidemiology Biomarkers & Prevention, 20*(4), 658-664. doi:10.1158/1055-9965.EPI-10-1008

Chen, J., Wu, X., Pande, M., Amos, C. I., Killary, A. M., Sen, S., et al. (2011). Susceptibility locus for lung cancer at 15q25.1 is not associated with risk of pancreatic cancer. *Pancreas, 40*(6), 872-875. doi:10.1097/MPA.0b013e318219dafe

Chen, L. S., Johnson, E. O., Breslau, N., Hatsukami, D., Saccone, N. L., Grucza, R. A., et al. (2009). Interplay of genetic risk factors and parent monitoring in risk for nicotine dependence. *Addiction, 104*(10), 1731-1740. doi:10.1111/j.1360-0443.2009.02697.x

Chen, X., Chen, J., Williamson, V. S., An, S. S., Hettema, J. M., Aggen, S. H., et al. (2009). Variants in nicotinic acetylcholine receptors α5 and α3 increase risks to nicotine dependence. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 150*(7), 926-933. doi:10.1002/ajmg.b.30919

Chen, X., Gorlov, I. P., Merriman, K. W., Weng, S. F., Foy, M., Keener, G., et al. (2011). Association of smoking with tumor size at diagnosis in non-small cell lung cancer. *Lung Cancer, 74*(3), 378-383. doi:10.1016/j.lungcan.2011.04.020

Chilcoat, H. D., Dishion, T. J., & Anthony, J. C. (1995). Parent monitoring and the incidence of drug sampling in urban elementary school children. *American Journal of Epidemiology, 141*(1), 25-31.

Cohen, S., Lichtenstein, E., Prochaska, J. O., Rossi, J. S., Gritz, E. R., Carr, C. R., et al. (1989). Debunking myths about self-quitting. Evidence from 10 prospective studies of persons who attempt to quit smoking by themselves. *American Psychologist, 44*(11), 1355-1365.

Conti, D. V., Lee, W., Li, D., Liu, J., Van den Berg, D., Thomas, P. D., et al. (2008). Nicotinic acetylcholine receptor β2 subunit gene implicated in a systems-based candidate gene study of smoking cessation. *Human Molecular Genetics, 17*(18), 2834-2848. doi:10.1093/hmg/ddn181

Cox, L. S., Tiffany, S. T., & Christen, A. G. (2001). Evaluation of the brief questionnaire of smoking urges (QSU-brief) in laboratory and clinical settings. *Nicotine and Tobacco Research, 3*(1), 7-16. 10.1080/14622200020032051

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology, 32*(1), 1-22.

David, S. P., & Munafo, M. R. (2008). Genetic variation in the dopamine pathway and smoking cessation. *Pharmacogenomics, 9*(9), 1307-1321.

Dawn Teare, M., & Barrett, J. H. (2005). Genetic linkage studies. *Lancet, 366*(9490), 1036-1044. S0140-6736(05)67382-5 [pii]

10.1016/S0140-6736(05)67382-5

De Ruyck, K., Nackaerts, K., Beels, L., Werbrouck, J., De Volder, A., Meysman, M., et al. (2010). Genetic variation in three candidate genes and nicotine dependence, withdrawal and smoking cessation in hospitalized patients. *Pharmacogenomics, 11*(8), 1053-1063. doi:10.2217/pgs.10.75

Decker, M. W., Sullivan, J. P., Arneric, S. P., & Williams, M. (2000). Neuronal nicotinic receptors: Novel targets for CNS therapeutics. *Neuropharmacology*.

Deeks, J., Higgins, J., & Altman, D. (Eds.). (2011). *Chapter 9: Analysing data and undertaking meta-analyses.* (Version 5.1.0 (updated March 2011) ed.): The Cochrane Collaboration, 2011.

Dick, D. M., Viken, R., Purcell, S., Kaprio, J., Pulkkinen, L., & Rose, R. J. (2007). Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. *Journal of Abnormal Psychology, 116*(1), 213-218. doi:10.1037/0021-843X.116.1.213

Doll, R. (1998). Uncovering the effects of smoking: historical perspective. *Statistical Methods in Medical Research, 7*, 87-117.

Doll, R., & Hill, A. B. (1950). Smoking and carcinoma of the lung; preliminary report. *British Medical Journal, 2*(4682), 739-748.

Doll, R., & Hill, A. B. (1954). The mortality of doctors in relation to their smoking habits; a preliminary report. *British Medical Journal, 1*(4877), 1451-1455.

Doll, R., & Hill, A. B. (1956). Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *British Medical Journal, 2*(5001), 1071-1081.

Donny, E. C., Caggiula, A. R., Knopf, S., & Brown, C. (1995). Nicotine self-administration in rats. *Psychopharmacology (Berl), 122*(4), 390-394.

Ducci, F., Kaakinen, M., Pouta, A., Hartikainen, A. L., Veijola, J., Isohanni, M., et al. (2011). TTC12-ANKK1-DRD2 and CHRNA5-CHRNA3-CHRNB4 influence different pathways leading to smoking behavior from adolescence to mid-adulthood. *Biological Psychiatry, 69*(7), 650-660. doi:10.1016/j.biopsych.2010.09.055

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. doi: 10.1111/j.0006-341X.2000.00455.x

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634. Retrieved from http://www.bmj.com/

Egger, M., & Smith, G. D. (1997). Meta-Analysis. Potentials and promise. *BMJ, 315*(7119), 1371-1374.

Egger, M., Smith, G. D., & Phillips, A. N. (1997). Meta-analysis: principles and procedures. *British Medical Journal, 315*(7121), 1533-1537.

English, J. P., Willius, F. A., & Berkson, J. (1940). Tobacco and coronary disease. *Journal of the American Medical Association, 115*(16), 1327-1329.

Erlich, P. M., Hoffman, S. N., Rukstalis, M., Han, J. J., Chu, X., Linda Kao, W. H., et al. (2010). Nicotinic acetylcholine receptor genes on chromosome 15q25.1 are associated with nicotine and opioid dependence severity. *Human Genetics, 128*(5), 491-499. doi:10.1007/s00439-010-0876-6

Etter, J. F., Hoda, J. C., Perroud, N., Munafo, M., Buresi, C., Duret, C., et al. (2009). Association of genes coding for the α-4, α-5, β-2 and β-3 subunits of nicotinic receptors with cigarette smoking and nicotine dependence. *Addictive Behaviors, 34*(9), 772-775. doi:10.1016/j.addbeh.2009.05.010

Flint, J., & Munafo, M. R. (2008). Forum: Interactions between gene and environment. *Current Opinion in Psychiatry, 21*(4), 315-317. doi:10.1097/YCO.0b013e328306a791

Fowler, C. D., Lu, Q., Johnson, P. M., Marks, M. J., & Kenny, P. J. (2011). Habenular alpha5 nicotinic receptor subunit signalling controls nicotine intake. *Nature, 471*(7340), 597-601. doi:10.1038/nature09797

Fowler, T., Lifford, K., Shelton, K., Rice, F., Thapar, A., Neale, M. C., et al. (2007). Exploring the relationship between genetic and environmental influences on initiation and progression of substance use. *Addiction, 102*(3), 413-422. doi:10.1111/j.1360-0443.2006.01694.x

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., et al. (2012). Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*. dys066 [pii]
10.1093/ije/dys066

Freathy, R. M., Kazeem, G. R., Morris, R. W., Johnson, P. C., Paternoster, L., Ebrahim, S., et al. (2011). Genetic variation at CHRNA5-CHRNA3-CHRNB4 interacts with smoking status to influence body mass index. *International Journal of Epidemiology, 40*(6), 1617-1628. dyr077 [pii]
10.1093/ije/dyr077

Freathy, R. M., Ring, S. M., Shields, B., Galobardes, B., Knight, B., Weedon, M. N., et al. (2009). A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) is associated with a reduced ability of women to quit smoking in pregnancy. *Human Molecular Genetics, 18*(15), 2922-2927. doi:10.1093/hmg/ddp216

Furberg, H., Kim, Y., Dackor, J., Boerwinkle, E., Franceschini, N., Ardissino, D., et al. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics, 42*(5), 441-447. doi:10.1038/ng.571

Gago-Dominguez, M., Jiang, X., Conti, D. V., Castelao, J. E., Stern, M. C., Cortessis, V. K., et al. (2011). Genetic variations on chromosomes 5p15 and 15q25 and bladder cancer risk: findings from the Los Angeles-Shanghai bladder case-control study. *Carcinogenesis, 32*(2), 197-202. doi:10.1093/carcin/bgq233

Girard, N., Lou, E., Azzoli, C. G., Reddy, R., Robson, M., Harlan, M., et al. (2010). Analysis of genetic variants in never-smokers with lung cancer facilitated by an internet-based blood collection protocol: A preliminary report. *Clinical Cancer Research, 16*(2), 755-763. doi:10.1158/1078-0432.CCR-09-2437

Golding, J., Pembrey, M., & Jones, R. (2001). ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatric and Perinatal Epidemiology, 15*(1), 74-87.

Gotti, C., Zoli, M., & Clementi, F. (2006). Brain nicotinic acetylcholine receptors: native subtypes and their relevance. *Trends in Pharmacological Sciences, 27*(9), 482-491. doi:10.1016/j.tips.2006.07.004

Gourlay, S. G., Stead, L. F., & Benowitz, N. L. (2004). Clonidine for smoking cessation. *Cochrane Database Syst Rev*(3), CD000058. 10.1002/14651858.CD000058.pub2

Gratacos, M., Gonzalez, J. R., Mercader, J. M., de Cid, R., Urretavizcaya, M., & Estivill, X. (2007). Brain-derived neurotrophic factor Val66Met and psychiatric disorders: meta-analysis of case-control studies confirm association to substance-related disorders, eating disorders, and schizophrenia. *Biological Psychiatry, 61*(7), 911-922. doi:10.1016/j.biopsych.2006.08.025

Greenbaum, L., Kanyas, K., Karni, O., Merbl, Y., Olender, T., Horowitz, A., et al. (2006). Why do young women smoke? I. Direct and interactive effects of environment, psychological characteristics and nicotinic cholinergic receptor genes. *Molecular Psychiatry, 11*(3), 312-322. doi:10.1038/sj.mp.4001774

Greenbaum, L., Rigbi, A., Teltsh, O., & Lerer, B. (2009). Role of genetic variants in the CHRNA5-CHRNA3-CHRNB4 cluster in nicotine dependence risk: Importance of gene-environment interplay. *Molecular Psychiatry, 14*(9), 828-830. doi:10.1038/mp.2009.25

Grucza, R. A., Johnson, E. O., Krueger, R. F., Breslau, N., Saccone, N. L., Chen, L. S., et al. (2010). Incorporating age at onset of smoking into genetic models for nicotine dependence: evidence for interaction with multiple genes. *Addiction Biology, 15*(3), 346-357. doi:10.1111/j.1369-1600.2010.00220.x

Grucza, R. A., Wang, J. C., Stitzel, J. A., Hinrichs, A. L., Saccone, S. F., Saccone, N. L., et al. (2008). A risk allele for nicotine dependence in CHRNA5 is a protective allele for cocaine dependence. *Biological Psychiatry, 64*(11), 922-929. doi:10.1016/j.biopsych.2008.04.018

Guillem, K., Vouillac, C., Azar, M. R., Parsons, L. H., Koob, G. F., Cador, M., et al. (2005). Monoamine oxidase inhibition dramatically increases the motivation to self-administer nicotine in rats. *Journal of Neuroscience, 25*(38), 8593-8600. 25/38/8593 [pii]
10.1523/JNEUROSCI.2139-05.2005

Hansen, H. M., Xiao, Y., Rice, T., Bracci, P. M., Wrensch, M. R., Sison, J. D., et al. (2010). Fine mapping of chromosome 15q25.1 lung cancer susceptibility in African-Americans. *Human Molecular Genetics, 19*(18), 3652-3661. doi:10.1093/hmg/ddq268

Hatsukami, D. K., Perkins, K. A., Lesage, M. G., Ashley, D. L., Henningfield, J. E., Benowitz, N. L., et al. (2010). Nicotine reduction revisited: science and future directions. *Tobacco Control, 19*(5), e1-10. 19/5/e1 [pii]
10.1136/tc.2009.035584

Health, A. o. S. a. (2012). *Smoking statistics: Who smokes and how much?*

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. O. (1991). The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction, 86*(9), 1119-1127.

Heishman, S. J., Kleykamp, B. A., & Singleton, E. G. (2010). Meta-analysis of the acute effects of nicotine and smoking on human performance. *Psychopharmacology (Berl), 210*(4), 453-469. 10.1007/s00213-010-1848-1

Henningfield, J. E., Rose, C. A., & Zeller, M. (2006). Tobacco industry litigation position on addiction: continued dependence on past views. *Tobacco Control, 15 Suppl 4*, iv27-36. 15/suppl_4/iv27 [pii]

10.1136/tc.2005.013789

Henningfield, J. E., & Zeller, M. (2002). Could science-based regulation make tobacco products less addictive? *Yale Journal of Health Policy, Law and Ethics, 3*(1), 127-138.

Heron, J., Hickman, M., Macleod, J., & Munafo, M. R. (2011). Characterizing patterns of smoking initiation in adolescence: comparison of methods for dealing with missing data. *Nicotine and Tobacco Research, 13*(12), 1266-1275. doi:10.1093/ntr/ntr161

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-560. 10.1136/bmj.327.7414.557

327/7414/557 [pii]

Hong, L. E., Yang, X., Wonodi, I., Hodgkinson, C. A., Goldman, D., Stine, O. C., et al. (2011). A CHRNA5 allele related to nicotine addiction and schizophrenia. *Genes, Brain and Behavior, 10*(5), 530-535. doi:10.1111/j.1601-183X.2011.00689.x

Hughes, J. R., Gulliver, S. B., Fenwick, J. W., Valliere, W. A., Cruser, K., Pepper, S., et al. (1992). Smoking cessation among self-quitters. *Health Psychology, 11*(5), 331-334.

Hung, R. J., McKay, J. D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., et al. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature, 452*(7187), 633-637. doi:10.1038/nature06885

Hurt, R. D., & Robertson, C. R. (1998). Prying open the door to the tobacco industry's secrets about nicotine: the Minnesota Tobacco Trial. *Journal of the American Medical Association, 280*(13), 1173-1181. jlm80013 [pii]

Improgo, M. R., Scofield, M. D., Tapper, A. R., & Gardner, P. D. (2010). The nicotinic acetylcholine receptor CHRNA5/A3/B4 gene cluster: dual role in nicotine addiction and lung cancer. *Prog Neurobiol, 92*(2), 212-226. S0301-0082(10)00112-7 [pii]

10.1016/j.pneurobio.2010.05.003

Jackson, K. J., Marks, M. J., Vann, R. E., Chen, X., Gamage, T. F., Warner, J. A., et al. (2010). Role of alpha5 nicotinic acetylcholine receptors in pharmacological and behavioral effects of nicotine in mice. *Journal of Pharmacology and Experimental Therapeutics, 334*(1), 137-146. doi:10.1124/jpet.110.165738

Janes, A. C., Smoller, J. W., David, S. P., Frederick, B. D., Haddad, S., Basu, A., et al. (2011). Association between CHRNA5 genetic variation at rs16969968 and brain reactivity to smoking images in nicotine dependent women. *Drug and Alcohol Dependence*. doi:10.1016/j.drugalcdep.2011.06.009

Jaworowska, E., Trubicka, J., Lener, M. R., Masojc, B., Zlowocka-Perlowska, E., McKay, J. D., et al. (2011). Smoking related cancers and loci at chromosomes 15q25, 5p15, 6p22.1 and 6p21.33 in the Polish population. *PLoS ONE, 6*(9), e25057. doi:10.1371/journal.pone.0025057

Johnson, E. O., Chen, L. S., Breslau, N., Hatsukami, D., Robbins, T., Saccone, N. L., et al. (2010). Peer smoking and the nicotinic receptor genes: an examination of genetic and environmental risks for nicotine dependence. *Addiction, 105*(11), 2014-2022. doi:10.1111/j.1360-0443.2010.03074.x

Jones, R. W., Ring, S., Tyfield, L., Hamvas, R., Simmons, H., Pembrey, M., et al. (2000). A new human genetic resource: a DNA bank established as part of the Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC). *European Journal of Human Genetics, 8*(9), 653-660. doi:10.1038/sj.ejhg.5200502

Kaprio, J. (2009). Genetic epidemiology of smoking behavior and nicotine dependence. *COPD, 6*(4), 304-306. 10.1080/15412550903049165 [pii]

Kaur-Knudsen, D., Bojesen, S. E., Tybjaerg-Hansen, A., & Nordestgaard, B. G. (2011). Nicotinic acetylcholine receptor polymorphism, smoking behavior, and tobacco-related cancer and lung and cardiovascular diseases: A cohort study. *Journal of Clinical Oncology, 29*(21), 2875-2882. doi:10.1200/JCO.2010.32.9870

Kendler, K. S., Chen, X., Dick, D., Maes, H., Gillespie, N., Neale, M. C., et al. (2012). Recent advances in the genetic epidemiology and molecular genetics of substance use disorders. *Nature Neuroscience, 15*(2), 181-189. nn.3018 [pii]

10.1038/nn.3018

Kendler, K. S., Neale, M. C., Sullivan, P., Corey, L. A., Gardner, C. O., & Prescott, C. A. (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychological Medicine, 29*(2), 299-308. doi:10.1017/S0033291798008022

Keskitalo, K., Broms, U., Heloidie, vaara, M., Ripatti, S., Surakka, I., et al. (2009). Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. *Human Molecular Genetics, 18*(20), 4007-4012. doi:10.1093/hmg/ddp322

Kim, D. K., Hersh, C. P., Washko, G. R., Hokanson, J. E., Lynch, D. A., Newell, J. D., et al. (2011). Epidemiology, radiology, and genetics of nicotine dependence in COPD. *Respiratory Research, 12*(1), 9. doi:10.1186/1465-9921-12-9

Kirkwood, B. R., & Sterne, J. A. (2003). *Essential Medical Statistics*: Blackwell Science.

Laird, R. D., Pettit, G. S., Bates, J. E., & Dodge, K. A. (2003). Parents' monitoring-relevant knowledge and adolescents' delinquent behavior: evidence of correlated developmental changes and reciprocal influences. *Child Development, 74*(3), 752-768. doi:10.1111/1467-8624.00566

Lambrechts, D., Buysschaert, I., Zanen, P., Coolen, J., Lays, N., Cuppens, H., et al. (2010). The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. *American Journal of Respiratory and Critical Care Medicine, 181*(5), 486-493. doi:10.1164/rccm.200909-1364OC

Landi, M. T., Chatterjee, N., Yu, K., Goldin, L. R., Goldstein, A. M., Rotunno, M., et al. (2009). A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *American Journal of Human Genetics, 85*(5), 679-691. doi:10.1016/j.ajhg.2009.09.012

Lando, H. A., Hipple, B. J., Muramoto, M., Klein, J. D., Prokhorov, A. V., Ossip, D. J., et al. (2010). Tobacco is a global paediatric concern. *Bulletin of the World Health Organization, 88*(1), 2. doi:10.2471/BLT.09.069583

Le Foll, B., & Goldberg, S. R. (2009). Effects of nicotine in experimental animals and humans: an update on addictive properties. *Handbook of Experimental Pharmacology*(192), 335-367. 10.1007/978-3-540-69248-5_12

Le Marchand, L., Derby, K. S., Murphy, S. E., Hecht, S. S., Hatsukami, D., Carmella, S. G., et al. (2008). Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Research, 68*(22), 9137-9140. doi:10.1158/0008-5472.CAN-08-2271

Leal, S. M. (2005). Detection of genotyping errors and pseudo-SNPs via deviations from Hardy-Weinberg equilibrium. *Genet Epidemiol, 29*(3), 204-214. 10.1002/gepi.20086

Lessov, C. N., Martin, N. G., Statham, D. J., Todorov, A. A., Slutske, W. S., Bucholz, K. K., et al. (2004). Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychological Medicine, 34*(5), 865-879. doi:10.1017/S0033291703001582

Li, M. D., Cheng, R., Ma, J. Z., & Swan, G. E. (2003). A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction, 98*(1), 23-31. doi:10.1046/j.1360-0443.2003.00295.x

Lickint, F. (1929). Tabak und tabakrauch als ätiologischer factor des carcinoms. *Zeitschr. Krebsforsch, 30*, 349-365.

Lips, E. H., Gaborieau, V., McKay, J. D., Chabrier, A., Hung, R. J., Boffetta, P., et al. (2009). Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *International Journal of Epidemiology, 39*(2), 563-577. doi:10.1093/ije/dyp288

Lips, E. H., Gaborieau, V., McKay, J. D., Chabrier, A., Hung, R. J., Boffetta, P., et al. (2010). Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *International Journal of Epidemiology, 39*(2), 563-577. doi:10.1093/ije/dyp288

Little, J., Higgins, J. P., Ioannidis, J. P., Moher, D., Gagnon, F., von Elm, E., et al. (2009). STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS Med, 6*(2), e22. 08-PLME-GG-0997 [pii]
10.1371/journal.pmed.1000022

Liu, J. Z., Tozzi, F., Waterworth, D. M., Pillai, S. G., Muglia, P., Middleton, L., et al. (2010). Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature Genetics, 42*(5), 436-440. doi:10.1038/ng.572

Liu, P., Vikis, H. G., Wang, D., Lu, Y., Wang, Y., Schwartz, A. G., et al. (2008). Familial aggregation of common sequence variants on 15q24-25.1 in lung cancer. *Journal of the National Cancer Institute, 100*(18), 1326-1330. doi:10.1093/jnci/djn268

Liu, Q. R., Drgon, T., Johnson, C., Walther, D., Hess, J., & Uhl, G. R. (2006). Addiction molecular genetics: 639,401 SNP whole genome association

identifies many "cell adhesion" genes. *Am J Med Genet B Neuropsychiatr Genet, 141B*(8), 918-925. 10.1002/ajmg.b.30436

Lombard, H. L., & Doering, C. R. (1928). Cancer Studies in Massachusetts: Habits, characteristics and environment of individuals with and without cancer. *New England Journal of Medicine, 198*(10), 481-487.

Lopez, A. D., Collishaw, N. E., & Piha, T. (1994). A descriptive model of the cigarette epidemic in developed countries. *Tobacco Control, 3*, 242-247.

Lori, A., Tang, Y., O'Malley, S., Picciotto, M. R., Wu, R., Conneely, K. N., et al. (2011). The galanin receptor 1 gene associates with tobacco craving in smokers seeking cessation treatment. *Neuropsychopharmacology, 36*(7), 1412-1420. doi:10.1038/npp.2011.25

Maes, H. H., Neale, M. C., Chen, X., Chen, J., Prescott, C. A., & Kendler, K. S. (2011). A twin association study of nicotine dependence with markers in the CHRNA3 and CHRNA5 genes. *Behavior Genetics, 41*(5), 680-690. doi:10.1007/s10519-011-9476-z

Marques-Vidal, P., Kutalik, Z., Paccaud, F., Bergmann, S., Waeber, G., Vollenweider, P., et al. (2011). Variant within the promoter region of the CHRNA3 gene associated with FTN dependence is not related to self-reported willingness to quit smoking. *Nicotine & Tobacco Research, 13*(9), 833-839. doi:10.1093/ntr/ntr084

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine, 3*(11), 2011-2030. doi:10.1371/journal.pmed.0030442

McKay, J. D., Hung, R. J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G., et al. (2008). Lung cancer susceptibility locus at 5p15.33. *Nature Genetics, 40*(12), 1404-1406. doi:10.1038/ng.254

Morris, P. (1992). Philip Morris Behavioral Research Program. *Tobacco Industry Documents: Philip Morris; Bates Number: 2021423427-3461, 1992.* , from http://legacy.library.ucsf.edu/action/document/page?tid=lum88d00

Muller, F. H. (1939). Tabakmissbrauch und lungencarcinom. *Zeitschr. Krebsforsch, 49*, 57-85.

Munafo, M. R., Clark, T. G., & Flint, J. (2004). Assessing publication bias in genetic association studies: evidence from a recent meta-analysis. *Psychiatry Research, 129*(1), 39-44. doi:10.1016/j.psychres.2004.06.011

Munafo, M. R., & Flint, J. (2009). Replication and heterogeneity in gene x environment interaction studies. *Int J Neuropsychopharmacol, 12*(6), 727-729. S1461145709000479 [pii]
10.1017/S1461145709000479

Munafo, M. R., & Johnstone, E. C. (2008). Genes and cigarette smoking. *Addiction, 103*(6), 893-904. doi:10.1111/j.1360-0443.2007.02071.x

Munafo, M. R., Johnstone, E. C., Walther, D., Uhl, G. R., Murphy, M. F., & Aveyard, P. (2011). CHRNA3 rs1051730 genotype and short-term smoking cessation. *Nicotine & Tobacco Research, 13*(10), 982-988. doi:10.1093/ntr/ntr106

Munafo, M. R., Timofeeva, M. N., Morris, R. W., Prieto-Merino, D., Sattar, N., Brennan, P., et al. (2012). Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *Journal of the National Cancer Institute, 104*(10), 740-748. doi:10.1093/jnci/djs191

NCI. (2009). Phenotypes and Endophenotypes: Foundations for Genetic Studies of Nicotine Use and Dependence. Tobacco Control Monograph No. 20. In N. I. o. H. U.S. Department of Health and Human Services, National Cancer Institute (Ed.) (Vol. NIH Publication No. 09-6366): Bethesda, MD.

Paterson, D., & Nordberg, A. (2000). Neuronal nicotinic receptors in the human brain. *Prog Neurobiol, 61*(1), 75-111. S0301-0082(99)00045-3 [pii]

Pearl, R. (1938). Tobacco smoking and longevity. *Science, 87*, 216-217.

Picciotto, M. R., Zoli, M., Rimondini, R., Lena, C., Marubio, L. M., Pich, E. M., et al. (1998). Acetylcholine receptors containing the beta2 subunit are involved in the reinforcing properties of nicotine. *Nature, 391*(6663), 173-177. doi:10.1038/34413

Piko, B. F., & Kovacs, E. (2010). Do parents and school matter? Protective factors for adolescent substance use. *Addictive Behaviors, 35*(1), 53-56. doi:10.1016/j.addbeh.2009.08.004

Pillai, S. G., Ge, D., Zhu, G., Kong, X., Shianna, K. V., Need, A. C., et al. (2009). A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genetics, 5*(3). doi:10.1371/journal.pgen.1000421

Plomin, R., DeFries, J., McClearn, G., & McGuffin, P. (2008). *Behavioral Genetics* (5 ed.). New York: Worth Publishers.

Pokhrel, P., Unger, J. B., Wagner, K. D., Ritt-Olson, A., & Sussman, S. (2008). Effects of parental monitoring, parent-child communication, and parents' expectation of the child's acculturation on the substance use behaviors of urban, Hispanic adolescents. *Journal of Ethnicity in Substance Abuse, 7*(2), 200-213. doi:10.1080/15332640802055665

Ray, R., Mitra, N., Baldwin, D., Guo, M., Patterson, F., Heitjan, D. F., et al. (2010). Convergent evidence that choline acetyltransferase gene variation is associated with prospective smoking cessation and nicotine dependence. *Neuropsychopharmacology, 35*(6), 1374-1382. doi:10.1038/npp.2010.7

Rigbi, A., Kanyas, K., Yakir, A., Greenbaum, L., Pollak, Y., Ben-Asher, E., et al. (2008). Why do young women smoke? V. Role of direct and interactive effects of nicotinic cholinergic receptor gene variation on neurocognitive function. *Genes, Brain and Behavior, 7*(2), 164-172. doi:10.1111/j.1601-183X.2007.00329.x

Riley, B. P. (2008). Commentary on 'The case for gene-environment interactions in psychiatry'. *Current Opinion in Psychiatry, 21*(4), 324-325. doi:10.1097/01.yco.0000320757.22733.30

Rodriguez, S., Cook, D. G., Gaunt, T. R., Nightingale, C. M., Whincup, P. H., & Day, I. N. (2011). Combined analysis of CHRNA5, CHRNA3 and CYP2A6 in relation to adolescent smoking behaviour. *Journal of Psychopharmacology, 25*(7), 915-923. doi:10.1177/0269881111405352

Rosenthal, R. (1991). Quality-weighting of studies in meta-analytic research. *Psychotherapy Research, 1*, 25-28.

Sacco, K. A., Bannon, K. L., & George, T. P. (2004). Nicotinic receptor mechanisms and cognition in normal states and neuropsychiatric disorders. *Journal of Psychopharmacology, 18*(4), 457-474. doi:10.1177/0269881104047273

Saccone, N. L., Culverhouse, R. C., Schwantes-An, T. H., Cannon, D. S., Chen, X., Cichon, S., et al. (2010a). Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer

and COPD. *PLoS Genetics, 6*(8), e1001053. doi:10.1371/journal.pgen.1001053

Saccone, N. L., Culverhouse, R. C., Schwantes-An, T. H., Cannon, D. S., Chen, X., Cichon, S., et al. (2010b). Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genetics, 6*(8). doi:10.1371/journal.pgen.1001053

Saccone, N. L., Saccone, S. F., Hinrichs, A. L., Stitzel, J. A., Duan, W., Pergadia, M. L., et al. (2009). Multiple distinct risk loci for nicotine dependence identified by dense coverage of the complete family of nicotinic receptor subunit (CHRN) genes. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 150*(4), 453-466. doi:10.1002/ajmg.b.30828

Saccone, N. L., Wang, J. C., Breslau, N., Johnson, E. O., Hatsukami, D., Saccone, S. F., et al. (2009). The CHRNA5-CHRNA3-CHRNB4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Research, 69*(17), 6848-6856. doi:10.1158/0008-5472.CAN-09-0786

Saccone, S. F., Hinrichs, A. L., Saccone, N. L., Chase, G. A., Konvicka, K., Madden, P. A. F., et al. (2007). Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics, 16*(1), 36-49. doi:10.1093/hmg/ddl438

Sakoda, L. C., Loomis, M. M., Doherty, J. A., Neuhouser, M. L., Barnett, M. J., Thornquist, M. D., et al. (2011). Chromosome 15q24-25.1 variants, diet, and lung cancer susceptibility in cigarette smokers. *Cancer Causes & Control, 22*(3), 449-461. doi:10.1007/s10552-010-9716-1

Sarginson, J. E., Killen, J. D., Lazzeroni, L. C., Fortmann, S. P., Ryan, H. S., Schatzberg, A. F., et al. (2011). Markers in the 15q24 nicotinic receptor subunit gene cluster (CHRNA5-A3-B4) predict severity of nicotine addiction and response to smoking cessation therapy. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 156B*(3), 275-284. doi:10.1002/ajmg.b.31155

SCENIHR. (2010). *Addictiveness and Attractiveness of Tobacco Additives*.

Schochet, T. L., Kelley, A. E., & Landry, C. F. (2004). Differential behavioral effects of nicotine exposure in adolescent and adult rats. *Psychopharmacology, 175*(3), 265-273. doi:10.1007/s00213-004-1831-9

Schwartz, A. G., Cote, M. L., Wenzlaff, A. S., Land, S., & Amos, C. I. (2009). Racial differences in the association between SNPs on 15q25.1, smoking behavior, and risk of non-small cell lung cancer. *Journal of Thoracic Oncology, 4*(10), 1195-1201. doi:10.1097/JTO.0b013e3181b244ef

Sen, S., & Burmeister, M. (2008). Hardy-Weinberg analysis of a large set of published association studies reveals genotyping error and a deficit of heterozygotes across multiple loci. *Hum Genomics, 3*(1), 36-52. J61886M812685171 [pii]

Sherva, R., Wilhelmsen, K., Pomerleau, C. S., Chasse, S. A., Rice, J. P., Snedecor, S. M., et al. (2008). Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit alpha 5 (CHRNA5) with smoking status and with 'pleasurable buzz' during early experimentation with smoking. *Addiction, 103*(9), 1544-1552. doi:10.1111/j.1360-0443.2008.02279.x

Shields, A., Lerman, C., & Sullivan, P. (2004). Translating emerging research on the genetics of smoking into clinical practice: ethical and social considerations.

Nicotine and Tobacco Research, 6(4), 675-688. 10.1080/14622200410001734058
HRPECNTAADWY6AG6 [pii]
Shiraishi, K., Kohno, T., Kunitoh, H., Watanabe, S. I., Goto, K., Nishiwaki, Y., et al. (2009). Contribution of nicotine acetylcholine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. *Carcinogenesis, 30*(1), 65-70. doi:10.1093/carcin/bgn257
Siedlinski, M., Cho, M. H., Bakke, P., Gulsvik, A., Lomas, D. A., Anderson, W., et al. (2011). Genome-wide association study of smoking behaviours in patients with COPD. *Thorax, 66*(10), 894-902. doi:10.1136/thoraxjnl-2011-200154
Sorice, R., Bione, S., Sansanelli, S., Ulivi, S., Athanasakis, E., Lanzara, C., et al. (2011). Association of a variant in the CHRNA5-A3-B4 gene cluster region to heavy smoking in the Italian population. *European Journal of Human Genetics, 19*(5), 593-596. doi:10.1038/ejhg.2010.240
Spitz, M. R., Amos, C. I., Dong, Q., Lin, J., & Wu, X. (2008). The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *Journal of the National Cancer Institute, 100*(21), 1552-1556. doi:10.1093/jnci/djn363
Statistics, O. f. N. (2012). *General Lifestyle Survey, 2010*.
Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53*(11), 1119-1129. S0895-4356(00)00242-0 [pii]
Stevens, V. L., Bierut, L. J., Talbot, J. T., Wang, J. C., Sun, J., Hinrichs, A. L., et al. (2008). Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiology Biomarkers and Prevention, 17*(12), 3517-3525. doi:10.1158/1055-9965.EPI-08-0585
Stolerman, I. P., & Jarvis, M. J. (1995). The scientific case that nicotine is addictive. *Psychopharmacology (Berl), 117*(1), 2-10; discussion 14-20.
Strasser, A. A., Lerman, C., Sanborn, P. M., Pickworth, W. B., & Feldman, E. A. (2007). New lower nicotine cigarettes can produce compensatory smoking and increased carbon monoxide exposure. *Drug and Alcohol Dependence, 86*(2-3), 294-300. doi:10.1016/j.drugalcdep.2006.06.017
Sullivan, P. F., & Kendler, K. S. (1999). The genetic epidemiology of smoking. *Nicotine and Tobacco Research, 1 Suppl 2*, S51-57. doi:10.1080/14622299050011811
The NHS Information Centre, L. S. (2011). *Statistics on Smoking: England, 2011*.
Thorgeirsson, T. E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K. P., et al. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature, 452*(7187), 638-642. doi:10.1038/nature06846
Thorgeirsson, T. E., Gudbjartsson, D. F., Surakka, I., Vink, J. M., Amin, N., Geller, F., et al. (2010). Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genetics, 42*(5), 448-453. doi:10.1038/ng.573
Thorgeirsson, T. E., & Stefansson, K. (2010). Commentary: gene-environment interactions and smoking-related cancers. *International Journal of Epidemiology, 39*(2), 577-579. doi: 10.1093/ije/dyp385

Tiffany, S. T., & Drobes, D. J. (1991). The development and initial validation of a questionnaire on smoking urges. *British Journal of Addiction, 86*(11), 1467-1476.

Timofeeva, M. N., McKay, J. D., Smith, G. D., Johansson, M., Byrnes, G. B., Chabrier, A., et al. (2011). Genetic polymorphisms in 15q25 and 19q13 loci, cotinine levels, and risk of lung cancer in EPIC. *Cancer Epidemiology Biomarkers & Prevention, 20*(10), 2250-2261. doi:10.1158/1055-9965.EPI-11-0496

Truong, T., Hung, R. J., Amos, C. I., Wu, X., Bickeboller, H., Rosenberger, A., et al. (2010). Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *Journal of the National Cancer Institute, 102*(13), 959-971. doi:10.1093/jnci/djq178

U.S. Department of Health and Human Services, C. f. D. C. a. P., National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. (2004). *The health consequences of smoking: a report of the Surgeon General*.

Uher, R. (2008). Forum: The case for gene-environment interactions in psychiatry. *Current Opinion in Psychiatry, 21*(4), 318-321. doi:10.1097/YCO.0b013e328306a7b9

Uhl, G. R., Drgon, T., Johnson, C., Fatusin, O. O., Liu, Q. R., Contoreggi, C., et al. (2008). "Higher order" addiction molecular genetics: convergent data from genome-wide association in humans and mice. *Biochem Pharmacol, 75*(1), 98-111. S0006-2952(07)00422-4 [pii]
10.1016/j.bcp.2007.06.042

USPHS. (1964). *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC.

Villegier, A. S., Salomon, L., Granon, S., Changeux, J. P., Belluzzi, J. D., Leslie, F. M., et al. (2006). Monoamine oxidase inhibitors allow locomotor and rewarding responses to nicotine. *Neuropsychopharmacology, 31*(8), 1704-1713. 1300987 [pii]
10.1038/sj.npp.1300987

Wall, J. D., & Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet, 4*(8), 587-597. 10.1038/nrg1123
nrg1123 [pii]

Wang, J., Spitz, M. R., Amos, C. I., Wilkinson, A. V., Wu, X., & Shete, S. (2010). Mediating effects of smoking and chronic obstructive pulmonary disease on the relation between the CHRNA5-A3 genetic locus and lung cancer risk. *Cancer, 116*(14), 3458-3462. doi:10.1002/cncr.25085

Wang, J. C., Grucza, R., Cruchaga, C., Hinrichs, A. L., Bertelsen, S., Budde, J. P., et al. (2009). Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. *Molecular Psychiatry, 14*(5), 501-510. doi:10.1038/mp.2008.42

Warburton, D. M. (1992). Nicotine as a cognitive enhancer. *Progress in Neuropsychopharmacology and Biological Psychiatry, 16*(2), 181-191. doi:10.1016/0278-5846(92)90069-Q

Ware, J., & Munafo, M. R. (2012). Testing times. *Addiction, 107*(12), 2077-2079. 10.1111/j.1360-0443.2012.03957.x

Ware, J. J., van den Bree, M. B., & Munafo, M. R. (2011). Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: A meta-analysis. *Nicotine and Tobacco Research, 13*(12), 1167-1175. doi:10.1093/ntr/ntr118

Wassenaar, C. A., Dong, Q., Wei, Q., Amos, C. I., Spitz, M. R., & Tyndale, R. F. (2011). Relationship between CYP2A6 and CHRNA5-CHRNA3-CHRNB4 variation and smoking behaviors and lung cancer risk. *Journal of the National Cancer Institute, 103*(17), 1342-1346. doi:10.1093/jnci/djr237

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Weiss, R. B., Baker, T. B., Cannon, D. S., Von Niederhausern, A., Dunn, D. M., Matsunami, N., et al. (2008). A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genetics, 4*(7). doi:10.1371/journal.pgen.1000125

WHO. (2008). *WHO Report on the Global Tobacco Epidemic, 2008: The MPOWER package*. Geneva: World Health Organisation. Retrieved from http://www.who.int/en/

Winterer, G., Mittelstrass, K., Giegling, I., Lamina, C., Fehr, C., Brenner, H., et al. (2010). Risk gene variants for nicotine dependence in the CHRNA5-CHRNA3-CHRNB4 cluster are associated with cognitive performance. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 153B*(8), 1448-1458. doi:10.1002/ajmg.b.31126

Wu, C., Hu, Z., Yu, D., Huang, L., Jin, G., Liang, J., et al. (2009). Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. *Cancer Research, 69*(12), 5065-5072. doi:10.1158/0008-5472.CAN-09-0081

Wyndor, E. L., & Graham, E. A. (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *Journal of the American Medical Association, 143*, 329-336.

Xie, P., Kranzler, H. R., Zhang, H., Oslin, D., Anton, R. F., Farrer, L. A., et al. (2011). Childhood adversity increases risk for nicotine dependence and interacts with alpha5 nicotinic acetylcholine receptor genotype specifically in males. *Neuropsychopharmacology*. doi:10.1038/npp.2011.240

Xun, W. W., Brennan, P., Tjonneland, A., Vogel, U., Overvad, K., Kaaks, R., et al. (2011). Single-nucleotide polymorphisms (5p15.33, 15q25.1, 6p22.1, 6q27 and 7p15.3) and lung cancer survival in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Mutagenesis, 26*(5), 657-666. doi:10.1093/mutage/ger030

Yang, P., Li, Y., Jiang, R., Cunningham, J. M., Zhang, F., & De Andrade, M. (2010). A rigorous and comprehensive validation: Common genetic variations and lung cancer. *Cancer Epidemiology Biomarkers and Prevention, 19*(1), 240-244. doi:10.1158/1055-9965.EPI-09-0710

Young, R. P., Hopkins, R. J., Hay, B. A., Epton, M. J., Black, P. N., & Gamble, G. D. (2008). Lung cancer gene associated with COPD: Triple whammy or possible confounding effect? *European Respiratory Journal, 32*(5), 1158-1164. doi:10.1183/09031936.00093908

Young, R. P., Hopkins, R. J., Hay, B. A., Epton, M. J., Mills, G. D., Black, P. N., et al. (2009). A gene-based risk score for lung cancer susceptibility in smokers and ex-smokers. *Postgraduate Medical Journal, 85*(1008), 515-524. doi:10.1136/pgmj.2008.077107

Zain, R. B., Anwar, A., Karen-Ng, L. P., Cheong, S. C., Zaini, Z., Prepagaran, N., et al. (2010). Selected transcriptome profiles of oral cancer suggestive of field cancerisation using second generation sequencing. *Oral Diseases, 16*, 521-521.

Zienolddiny, S., Skaug, V., Landvik, N. E., Ryberg, D., Phillips, D. H., Houlston, R., et al. (2009). The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. *Carcinogenesis, 30*(8), 1368-1371. doi:10.1093/carcin/bgp131