Published in final edited form as:

Curr Psychiatry Rep. 2014 October ; 16(10): 477. doi:10.1007/s11920-014-0477-5.

Determining the Causes and Consequences of Nicotine Dependence: Emerging Genetic Research Methods

Jennifer J. Ware, Ph.D.^{1,2} and Marcus R. Munafò, Ph.D.^{1,3}

¹MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, BS8 2BN, United Kingdom

²School of Social and Community Medicine, University of Bristol, Oakfield House, Bristol, BS8 2BN, United Kingdom

³UK Centre for Tobacco and Alcohol Studies and School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, United Kingdom

Abstract

Tobacco use remains the leading cause of preventable death worldwide. Establishing the genetic aetiology of tobacco use and dependence is an important first step in understanding the neurobiological mechanisms of tobacco use, and in turn the development of effective treatments. In addition, whilst the effects of tobacco use on a broad range of physical illnesses (e.g., lung cancer, respiratory disease, cardiovascular disease) are now well-established, the causal effects of tobacco use on a number of other outcomes remains to be established. Determining the causes and consequences of tobacco use therefore continues to be both a scientific and a public health priority. Here we review emerging methods in genetic research that allow stronger causal inferences to be drawn from observational data.

Keywords

Nicotine; Tobacco; Smoking; Dependence; Genetics; GWAS; Mendelian Randomisation; Recall-by-Genotype; Phenotype Refinement; Biomarkers

Introduction

The tobacco use epidemic continues to grow on a global scale, and remains the leading cause of preventable death worldwide [1, 2]. It has been recognised for some time that smoking behaviours are under a substantial degree of genetic influence [3, 4]. Identification of the specific genetic variants underlying these behaviours should clarify the neurobiological mechanisms of tobacco use, and is therefore an important first step towards: a) the development of novel treatments, and b) tailoring of existing pharmacotherapies. Whilst a decade of candidate gene association studies taught us relatively little about the genetic architecture of smoking phenotypes, genome-wide association studies (GWAS) have proven far more successful [5–7]. Preclinical studies, and follow-up studies in humans

employing carefully refined phenotypes, have further developed our understanding of these relationships. The identification of variants robustly associated with smoking behaviour has also allowed us to clarify the causal nature of association between tobacco use and co-occurring health outcomes, through employment of Mendelian randomisation techniques. Each of these genetic approaches will be discussed in turn, alongside new and emerging technologies.

GWAS, Imputation and Sequencing Technologies

The most robust finding to emerge from GWAS of smoking phenotypes is the association between the nicotinic receptor gene cluster *CHRNA5-A3-B4* on chromosome 15 (at 15q25) and smoking quantity. This gene cluster encodes three nicotinic acetylcholine receptor subunit proteins: $\alpha 5$, $\alpha 3$ and $\beta 4$. An association between rs16969968 in *CHRNA5* and nicotine dependence (ND) was first reported in 2007 in a candidate gene study [8], with the minor allele found to confer increased risk. The following year, the same locus (tagged by rs1051730 in *CHRNA3*, a variant highly correlated with rs16969968) was found to be associated with smoking quantity, this time identified in a GWAS conducted by Thorgeirsson and colleagues [9]. This study also demonstrated an association between rs1051730 and nicotine dependence and two smoking-related diseases, namely lung cancer and peripheral arterial disease, lending further weight to the finding. Notably, rs16969968 is a nonsynonymous variant, resulting in an amino acid change in the resultant $\alpha 5$ nicotinic receptor subunit protein. The association between this locus and smoking quantity has since been replicated in several consequent independent studies [5–7, 10, 11].

Genotype imputation in GWAS

Early GWAS of smoking behaviours typically examined associations with directly genotyped single nucleotide polymorphisms (SNPs) only, as determined from microarrays, which provide relatively sparse coverage of the genome, and include common SNPs only (i.e., minor allele frequency > 5%) [e.g., 9, 12]. In contrast, more recent GWAS now commonly employ imputation to expand genomic coverage, including capture of rare or low frequency variants. Imputation allows us to predict ('impute') genotypes that have not been directly genotyped on a microarray. In order to do so, microarray data is matched to a genome reference panel, which consists of densely genotyped (or sequenced) genomic data from multiple individuals. The genomic reference panels used for imputation are HapMap (<http://hapmap.ncbi.nlm.nih.gov/>) and 1000 Genomes (<http://www.1000genomes.org/>). The most recent release of 1000 Genomes (Phase I) provides a much larger set of SNPs (~37 million) relative to HapMap (~4 million). Imputation is used in GWAS for a number of reasons. Firstly, it can increase study power [13]. Secondly, it provides us with a more detailed view of particular genomic regions that contain many associated SNPs. This fine-mapping approach may allow us to identify actual causal variants not captured by microarrays. Finally, imputation allows us to combine GWAS results across studies which have used different GWAS chips, thus enabling meta-analysis [13].

Imputation has been used to great success in recent GWAS of smoking phenotypes, enabling the running of very large ($N > 75,000$), consortia-based GWAS meta-analyses, which have

identified a number of new genomic regions [5–7]. In addition to the *CHRNA5-A3-B4* region, genomic regions on chromosomes 8, 10 and 19 have been identified in relation to smoking quantity [5, 7]. In particular, two genes with clear biological relevance were identified – *CHRNA3/CHRNA6* on chromosome 8 and *CYP2A6* on chromosome 19 [5]. *CYP2A6* is the enzyme principally responsible for the metabolic inactivation of nicotine to cotinine [14, 15]. These studies also highlighted the importance of *BDNF* (chromosome 11) in relation to smoking initiation, and *DBH* (chromosome 9) in relation to smoking cessation [7]. Further, imputation has enabled fine-mapping of the 15q25 locus [6]. Liu et al. [6] used 1000 Genomes Pilot 1 data to fine-map the 15q25 region for association with smoking quantity. This allowed for analysis of virtually all common SNPs (MAF > 5%) in the region, and offered a five-fold increase in marker density compared to HapMap 2. Using this imputation approach, combined with meta-analysis, they identified variant rs55853698 as the variant with strongest evidence for association with smoking quantity at this locus. This variant, which is notably absent from HapMap2, is located within a promoter region of *CHRNA5*, and is a plausible candidate for affecting mRNA transcription. Conditioning on variant rs55853698 revealed a second independent signal within this region (rs6495308, located within *CHRNA3*). Conditioning on both SNPs left no residual signal, suggesting that these two variants could explain the full signal at 15q25 in relation to smoking quantity [6].

GWAS has certainly advanced our understanding of the genetic underpinnings of smoking initiation, cessation, and smoking quantity. However, despite these successes, the total proportion of variance explained by all variants identified through GWAS to date is far less than the heritability estimates indicated by earlier twin studies. This may be because the individual effects of SNPs studied are too small to reach the strict thresholds for genome-wide significance. It is also possible that rare variants (MAF < 1%), which cannot be very accurately imputed using current genomic reference panels (given the limited numbers of haplotypes on which they are comprised), may explain much of this missing heritability. It is also possible that the heritability values from twin studies are over-estimated. Finally, epistasis (multiplicative interaction among risk alleles) may explain additional fractions of the missing heritability.

Genome Sequencing

In contrast to GWAS, genome sequencing, involving the use of high-throughput sequencing technologies to determine the order of nucleotide bases present in a particular individual, provides high quality rare variant genotype calls. Up until very recently, the cost of this technology has prevented its widespread use (the first sequenced human genome, completed in 2003, cost ~\$3 billion); however, recent advances in technology have resulted in a dramatic drop in cost in recent years, and the \$1,000 genome is now a reality [16]. Sequencing techniques are now being employed to further our understanding of tobacco dependence and other smoking-related phenotypes. An example of such a project is the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN; <https://gscan.sph.umich.edu/>), led by Scott Vrieze at the University of Michigan. This is a collaborative effort comprising around 30 individual studies, with the aim of identifying genes associated with various nicotine- (and alcohol-) related phenotypes, including cigarettes per day, pack years, smoking initiation, and age of smoking initiation. The

GSCAN project comprises three complimentary approaches: GWAS meta-analysis, exome meta-analysis, and whole genome sequencing meta-analysis. All contributing studies will be imputed to a new haplotype reference panel currently under development, expected to have over 25,000 sequenced individuals, which will not only allow for high-density imputation, but also accurate imputation of low frequency variants. Further, given the size of the collaboration effort, the sample size promises to be extremely large, affording increased power to detect small genetic effects. The exome meta-analysis will in principle allow for the identification of rare, non-synonymous, exonic variants associated with nicotine-related phenotypes. Finally, the sequencing arm of the GSCAN project will be focused on following up variants identified through GWAS analyses, allowing evaluation at greatly increased genomic resolution. The complimentary use of GWAS and sequencing techniques has recently been used to great success [17]. In short, this large, collaborative project holds the potential to substantially advance our understanding of the genetic basis of smoking-related behaviours.

Phenotype refinement

Whilst GWAS, exome- and whole-genome sequencing techniques allow for the identification of genetic variants associated with specific phenotypes, they are not without limitations. Firstly, they often rely upon self-report measures of smoking phenotypes, frequently determined retrospectively. Further, assessments are often inconsistent across different studies. This combination of self-report, retrospective recall, and the need for phenotype harmonisation, may introduce considerable measurement error. Secondly, these approaches tell us little about the fundamental mechanism of action linking the identified variant to a specific behaviour. Phenotype refinement not only provides a means by which to understand these mechanisms, but also theoretically affords a ‘cleaner’ genetic signal, improving likelihood of effect replication.

Animal models

One means by which to explore the mechanistic impact of a particular gene on a specific phenotype involves the use of ‘knockout’ mouse models, involving the selective disabling of a target gene. The development of α_5 knockout mouse models has illustrated the role that the *CHRNA5* gene plays in determining response to nicotine. Such mice are considered a model of individuals with reduced *CHRNA5* gene function (i.e., carriers of the rs16969968 minor allele, associated with increased heaviness of smoking) [see 18 for description of previous in vitro work]. Using a nicotine self-administration paradigm, Fowler and colleagues [19] observed that α_5 knockout mice responded far more vigorously than wild-type mice (i.e., mice with no inactivated genes) for nicotine infusions at high doses. Whilst wild-type mice self-titrated delivery of nicotine dose to achieve a consistent, desired level during each test session (~1.5 mg/kg), knockout mice did not, consuming greater amounts as dosage increased. This seminal paper concluded that deficient α_5 signalling, specifically limited to the habenula/interpeduncular nucleus pathway, attenuates the negative effects of nicotine that normally serve to limit intake [19].

Candidate gene follow-up studies

Genetic variants identified through conventional GWAS can be followed up using a candidate gene approach, exploring the genetic variant so identified with greater phenotype precision. The utility of this approach has been demonstrated in relation to the observed 15q25 locus and self-reported smoking quantity association. Adopting this approach, Keskitalo and colleagues [20] examined the association of rs1051730/rs16969968 with self-reported smoking quantity and circulating levels of cotinine. Cotinine is the primary metabolite of nicotine, and an excellent biomarker of smoking quantity. They found that rs1051730/rs16969968 accounts for ~1% of the variation in self-reported smoking quantity, compared to ~4% of the variation in cotinine levels. Further extending this work, Munafò and colleagues [21] examined the same associations of rs1051730 with self-reported smoking quantity and circulating levels of cotinine in a much larger sample. In line with previous findings, they observed a much stronger relationship between rs1051730 and cotinine relative to self-reported daily cigarette consumption. Importantly, the rs1051730 and cotinine relationship was robust to adjustments made for self-reported smoking quantity. This suggests that even amongst *equal* cigarette consumers, there is still genetically-influenced variation in total nicotine exposure. Smaller, previous studies provide further support for this finding. Le Marchand and colleagues [22] examined the association between the same locus and total nicotine equivalents (an extremely accurate measure of nicotine exposure, encompassing nicotine, cotinine, related glucuronides, and 3-hydroxycotinine). The relationship observed was also robust for adjustments made for self-reported daily cigarette consumption. These observations are likely to be the result of differences in cigarette smoking style between smokers [21].

The key point illustrated by these studies is that refined and carefully characterised phenotypes, which afford greater precision of measurement, provide a 'cleaner' genetic signal, and improve likelihood of effect replication. Building upon this principle, we have recently conducted a GWAS meta-analysis of cotinine levels in daily smokers (Ware et al., in preparation). Combining GWAS with an objective assessment of a behavioural phenotype, using a relevant biomarker, affords improved statistical power and may help identify novel variants.

Recall-by-genotype studies

As illustrated in the preceding section, refining the nature of associations observed initially through GWAS can be achieved using follow-up candidate gene studies employing precisely assessed phenotypes. However, one limitation to this approach is that the measurement of detailed, precise phenotypes in large, adequately powered samples can be costly and time-consuming. One approach that circumvents these issues is the 'recall-by-genotype' method. The recall-by-genotype approach involves the selection of participants (or their biological samples) on the basis of a specific genotype, for further detailed, clinically-relevant, phenotype assessment. This approach is far more efficient than random sampling, enabling over-sampling of rare allele carriers, and enables the collection of extremely detailed phenotypic data that would be impractical to collect in a much larger sample for reasons of time and expense.

We are currently employing a recall-by-genotype approach to further explore the association between rs1051730/rs16969968 and smoking quantity [23]. As discussed above, the association observed between this locus and circulating cotinine is robust to adjustment for daily cigarette consumption. This suggests that genetically influenced variation in total nicotine exposure exists between *equal* cigarette consumers. Differences in cigarette smoking style (e.g., puff inhalation volumes, number of puffs taken per cigarette), termed ‘smoking topography’, likely underlie these differences [21]. It has previously been established that cigarette smokers modify their smoking behaviour to self-titrate nicotine to a level appropriate to individual need [24, 25]. What remains to be established however is whether these differences in smoking style mediate the relationship observed between rs1051730/rs16969968 and cotinine levels. To investigate this we are recalling current smokers from a large cohort study on the basis of minor and major homozygote status at rs1051730/rs16969968, assessing smoking topography and cotinine levels in both genotypic groups [23].

Mendelian randomisation

The identification of variants robustly associated with smoking behaviours (e.g., rs1051730/rs16969968 and smoking quantity) has not only developed our understanding of the aetiology of tobacco use, but has also allowed us to clarify the causal nature of association between tobacco use and co-occurring health- and socioeconomic-related outcomes, through employment of Mendelian randomisation (MR) techniques (see Figure 1).

Ascertaining causality between co-morbid conditions, such as tobacco use and depression, has proven problematic using conventional observational methodologies, given well-known problems of confounding and reverse causality. MR offers a solution to these difficulties [26]. MR involves the use of specific genetic variants that have been shown to be robustly associated with a specific exposure (e.g., smoking quantity); these are then used as proxies for said exposure. The principle of MR relies on the basic (but approximate) laws of Mendelian genetics (segregation and independent assortment). If these two laws hold then, at a population level, genetic variants will not be associated with the confounding factors that generally distort conventional observational studies. Further, the genes that individuals are born with should not be altered by environmental factors, which removes the issue of reverse causality. MR has been used to great success in illustrating the causal influence of tobacco use on a range of health outcomes, including offspring birth weight [27], body mass index [28, 29], and resting heart rate [29]. Large-scale collaborative efforts employing these techniques are currently underway to further our understanding of the causal impact of substance use on a host of other outcomes. The consortium for Causal Analysis Research in Tobacco and Alcohol (CARTA; <http://www.bris.ac.uk/expsych/research/brain/targ/research/collaborations/carta/>), established at the University of Bristol, is one such example. To date, this consortium comprises over 30 studies, spanning nine countries, with a total sample size in excess of 150,000. CARTA has completed five initial analyses, investigating the impact of cigarette smoking on depression and anxiety, regional adiposity, blood pressure and heart rate, serum vitamin D levels and income. The results of these analyses, which are currently in preparation, have demonstrated for example that smoking does not cause depression or

anxiety [30], but may result in higher levels of abdominal adiposity (as measured by waist circumference) for a given BMI [31].

Conclusions

The new and emerging genetic research methods discussed within this review have developed our understanding of not only the genetic architecture of smoking behaviours, but also our understanding of the causal impact of tobacco use on a spectrum of health outcomes. Large-scale GWAS efforts, employing high-density imputation, have identified several genes in relation to smoking quantity, initiation, and cessation. Consequent *in vitro*, preclinical, and follow-up candidate gene studies in humans have developed our understanding of the genetic mechanisms of action. New approaches, such as the recall-by-genotype method, may afford us further mechanistic insight. However, despite these successes, we can still only account for relatively little of the heritability of smoking behaviours. New sequencing approaches, which are increasingly becoming more affordable, hold promise and may offer the potential to discover rare or novel variants associated with smoking-related behaviours. Investigating the function of newly identified genetic variants must follow, through careful employment of the techniques discussed within this review (amongst others), which may ultimately lead to the development of novel treatments for tobacco dependence, and, as is the goal, an improvement in health outcomes.

Acknowledgements

Jennifer Ware is supported by a postdoctoral Research Fellowship from the Oak Foundation. Jennifer Ware and Marcus Munafò are members of the UK Centre for Tobacco and Alcohol Studies, a UK Clinical Research Council Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

References

Papers of particular interest, published recently, have been highlighted as:

* Of importance

** Of major importance

1. WHO. WHO report on the global tobacco epidemic, 2013. Geneva: World Health Organisation; 2013.
2. WHO. [cited 2014 20th May] Tobacco Fact Sheet. 2013. Available from: <http://www.who.int/mediacentre/factsheets/fs339/en/>
3. Carmelli D, et al. Genetic influence on smoking--a study of male twins. *N Engl J Med*. 1992; 327(12):829–33. [PubMed: 1508241]
4. Heath AC, et al. Genetic contribution to risk of smoking initiation: comparisons across birth cohorts and across cultures. *Journal of Substance Abuse*. 1993; 5(3):221–246. [PubMed: 8312729]
5. Thorgeirsson TE, et al. Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genetics*. 2010; 42(5):448–453. [PubMed: 20418888] [Extremely large consortia-based GWAS meta-analysis identifying CHRN3-CHRNA6 and CYP2A6 in relation to smoking quantity.]
6. Liu JZ, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature Genetics*. 2010; 42(5):436–440. [PubMed: 20418889]

7. Furberg H, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics*. 2010; 42(5):441–7. [PubMed: 20418890] [Large consortium-based GWAS meta-analysis of multiple smoking phenotypes, identifying DBH in relation to smoking cessation, BDNF in relation to smoking initiation, and CYP2A6 and CHRNA5-A3-B4 in relation to smoking quantity.]
8. Saccone SF, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics*. 2007; 16(1):36–49. [PubMed: 17135278]
9. Thorgeirsson TE, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008; 452(7187):638–642. [PubMed: 18385739]
10. David SP, et al. Genome-wide meta-analyses of smoking behaviors in African Americans. *Translational Psychiatry*. 2012; 2(5):e119. [PubMed: 22832964]
11. Ware JJ, van den Bree MB, Munafo MR. Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: A meta-analysis. *Nicotine and Tobacco Research*. 2011; 13(12):1167–1175. [PubMed: 22071378]
12. Caporaso N, et al. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One*. 2009; 4(2):e4653. [PubMed: 19247474]
13. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*. 2010; 11:499–511.
14. Tyndale RF, Sellers EM. Genetic variation in CYP2A6-mediated nicotine metabolism alters smoking behavior. *Ther Drug Monit*. 2002; 24(1):163–71. [PubMed: 11805739]
15. Hukkanen J, Jacob P 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. *Pharmacol Rev*. 2005; 57(1):79–115. [PubMed: 15734728]
16. Hayden EC. Is the \$1,000 genome for real? *Nature*. 2014
17. Holm H, et al. A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nat Genet*. 2011; 43(4):316–20. [PubMed: 21378987]
18. Bierut LJ, et al. Variants in the nicotinic receptors alter risk for nicotine dependence. *American Journal of Psychiatry*. 2008; 165(9):1163–1171. [PubMed: 18519524]
19. Fowler CD, et al. Habenular alpha5 nicotinic receptor subunit signalling controls nicotine intake. *Nature*. 2011; 471(7340):597–601. [PubMed: 21278726] [Preclinical research involving α_5 knockout models, indicating that deficient α_5 signalling, specifically limited to the habenula/interpeduncular nucleus pathway, attenuates the negative effects of nicotine that normally serve to limit intake.]
20. Keskitalo K, et al. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNA4) on chromosome 15. *Human Molecular Genetics*. 2009; 18(20):4007–4012. [PubMed: 19628476]
21. Munafo MR, et al. Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *Journal of the National Cancer Institute*. 2012; 104(10):740–8. [PubMed: 22534784] [Follow-up candidate gene study illustrating the improvement in statistical power offered by objective assessment of behavioural phenotypes relative to self-report measures.]
22. Le Marchand L, et al. Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Research*. 2008; 68(22):9137–9140. [PubMed: 19010884]
23. Ware JJ, et al. A recall-by-genotype study of CHRNA5-A3-B4 genotype, cotinine and smoking topography: study protocol. *BMC Medical Genetics*. 2014; 15(13)
24. McNeill A, Munafo MR. Reducing harm from tobacco use. *J Psychopharmacol*. 2013; 27(1):13–8. [PubMed: 23035032]
25. Strasser AA, et al. New lower nicotine cigarettes can produce compensatory smoking and increased carbon monoxide exposure. *Drug and Alcohol Dependence*. 2007; 86(2-3):294–300. [PubMed: 16930853]
26. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*. 2003; 32(1):1–22. [PubMed: 12689998] [Introduces a new technique enabling determination of

causation between variables in epidemiological datasets, using genetic variants as proxies for modifiable environmental factors.]

27. Tyrrell J, et al. Genetic variation in the 15q25 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNA4) interacts with maternal self-reported smoking status during pregnancy to influence birth weight. *Hum Mol Genet.* 2012; 21(24):5344–58. [PubMed: 22956269]
28. Freathy RM, et al. Genetic variation at CHRNA5-CHRNA3-CHRNA4 interacts with smoking status to influence body mass index. *International Journal of Epidemiology.* 2011; 40(6):1617–28. [PubMed: 21593077]
29. Asvold BO, et al. Causal associations of tobacco smoking with cardiovascular risk factors: a Mendelian randomization analysis of the HUNT Study in Norway. *Int J Epidemiol.* 2014
30. Taylor AE, et al. Investigating the possible causal association of smoking with depression and anxiety using Mendelian randomisation meta-analysis: The CARTA consortium. *BMJ Open.* under review.
31. Morris RW, et al. Investigating the possible causal association between smoking and regional adiposity using Mendelian randomisation meta-analysis: The CARTA consortium. in preparation.

Text Box 1. Glossary***Candidate gene association study:***

A study in which a specific genetic variant, selected on the basis of biological relevance to the phenotype of interest, is examined in relation to a phenotype. This approach is hypothesis-*driven*.

Genome-wide association study (GWAS):

A study in which genetic variation across the entire genome is examined in relation to a phenotype. This approach simultaneously examines the association of hundreds of thousands of genetic variants with a phenotype. This approach is hypothesis-*free*.

Genotype imputation:

Prediction of genotypes for genetic variants that have not been directly genotyped. This is achieved through computational algorithms that use directly genotyped variant information combined with use of a haplotype reference panel (e.g., HapMap) to impute this missing information.

Genome sequencing:

Determination of the order of all nucleotide bases present within the full genome of a particular individual.

Exome sequencing:

Determination of the order of nucleotide bases present within the exome (i.e., protein-coding region of the genome) of a particular individual.

Recall-by-genotype study:

A study in which participants, or their biological samples, are selected on the basis of a specific genotype, for detailed, clinically relevant, phenotype assessment.

Mendelian randomisation:

A form of instrumental variable analysis, in which a genetic variant robustly associated with a modifiable exposure is used as a proxy for said exposure, allowing causal inferences to be made about the relationship between exposure and a specific outcome. This approach reduces problems of confounding and reverse causality common to conventional observational methodologies.

Figure 1. Mendelian randomisation.

Mendelian randomisation is a form of instrumental variable analysis, in which a genetic variant robustly associated with a modifiable exposure (e.g., rs1051730 and smoking quantity) is used as a proxy for said exposure. Examination of the relationship between this variant and a specific outcome (e.g., depression), allows us to make *causal* inferences about the nature of the relationship between said exposure and outcome. Given the random assortment of alleles at time of gamete formation and conception, genetic variants should not be associated with confounding factors that can distort conventional observational studies. Further, individual genetic variation should not be influenced by environmental factors, which removes the issue of reverse causality.