A systems biology approach for the study of smoking and myocardial infarction

Tao Xu

München 2014

Eidesstattliche Versicherung

Xu, Tao

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

A systems biology approach for the study of smoking and myocardial infarction

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

19.03.2015, Munich

Ort, Datum

Unterschrift Doktorandin/Doktorand

Aus dem Helmhotz Zentrum München
Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)
Abteilung für Molecular Epidemiologie (AME)
Vorstand: Dr. Christian Gieger

A systems biology approach for the study of smoking and myocardial infarction

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität
München

vorgelegt von
Tao Xu

aus Wuxi
2015

Gedruckt mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuer: Prof. Dr. Thomas Illig

Zweitgutachter: Prof. Dr. Gerhard Scherer

Dekan: Prof. Dr. med. Dr. h. c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 09.03.2015

Table of Contents

List of Tables

List of Figures

Summary

Metabolomics has been proven to be a powerful tool to study complex phenotypes. It can present a snapshot of the current status of metabolism and provide a functional readout of the gene products. Complementing with other 'omics' techniques in systems biology studies, the integration of metabolomics with other 'omics', e.g. trancriptomics and epigenomics, will help to illustrate complex biological processes which are related to disease and environmental exposure.

This thesis presents three studies focusing on a lifestyle related environmental exposure—smoking and a disease which is related to the exposure—myocardial infarction (MI). The general aim of these studies is to establish links between smoking, intermediated biomarkers of disturbed metabolic pathways and MI. Establishment of how they are linked might enlarge our knowledge about the metabolic basis of these links.

The first study presented in this thesis aims to understand the effects of smoking and smoking cessation on human serum metabolite profile. Whilst smoking increases the risks of many diseases, including MI, the benefits of cessation is remarkable as it has shown to reduce the risk of MI in a very short time frame. The results presented in this thesis showed significant differences in metabolite profiles between current smokers, former smokers and never smokers. Amongst the 21 metabolites, which were found to be different between current smokers and never smokers, 19 were found reversible in former smokers. The results were furthermore confirmed in the prospective study of KORA S4→F4. Network analysis was applied to integrate smoking related genes and metabolites, which consistently showed the reversibility of the smoking effects on gene expression and metabolite profile. The reversibility of smoking related changes in serum metabolites also coincide with the reduced risk of

MI, which gives rise to the possibility of using these metabolites as potential biomarkers to characterize smoking related diseases.

Inspired by the first study, two other studies were initiated with different aims. The second study in this thesis aims to use multi-level 'omics' data to illustrate how smoking influences the metabolite profiles by alteration in DNA methylation and gene expression. Candidate biomarkers of smoking were first discovered separately in epigenomic, transcriptomic and metabolomic levels. Mediation analyzes were applied to assess the potential interactions between smoking, DNA methylation, gene expression and metabolites. In general, seven CpG sites showed significant mediation effects for the expression of the *LRRN3* gene. Amongst these seven, two were also significantly associated with the concentrations of LPC (18:2) and PC ae C34:3.

In the third study, three metabolites (arginine, LPC (17:0) and LPC (18:2)), which may serve as novel biomarkers for incident MI, were identified based on a targeted metabolomics approach in two prospective cohort studies. These metabolites significantly associated with MI in Cox regression models after adjustment for other MI risk factors, such as smoking and C-reactive protein (CRP). Inclusion of these metabolites in the established MI prediction models provided significant added predictive value. Additionally, the observation that these metabolites were associated with CRP indicates potential inflammatory process they are commonly involved in. Among the three metabolites listed above, arginine and LPC (18:2) are also associated with smoking as shown in the first study of this thesis, which implies the underlying metabolic relationships between smoking and MI.

In summary, this doctoral thesis reveals metabolites associated with smoking and MI. Using a systems biology approach, the effects of smoking on DNA methylation and

gene expression, which mediates the corresponding variations on metabolite concentrations, were analyzed by integrating multi-level ‘omics’ data. The metabolites associated with both smoking and MI may contribute to a deeper insight into the molecular basis between the link of MI and its risk factor—smoking.

Zusammenfassung

Metabolomics ist ein bewährtes und leistungsfähiges Hilfsmittel zur Erforschung von komplexen Phänotypen. Diese Technik bildet zum einen eine Momentaufnahme des gerade herrschenden Status des Metabolismus ab und dient zum anderen zum Auslesen von Genprodukten. Mit Hilfe des Zusammenspiels und der Einbindung anderer „-omics“ Technologien der Systembiologie können komplexe biologische Prozesse, die in Beziehung zu Krankheiten und Umwelteinflüssen stehen, dargestellt werden.

Diese Doktorarbeit beinhaltet drei Studien deren Fokus auf dem Thema Abhängigkeit der Lebensgewohnheiten von Umwelteinflüsse liegt - dem Rauchen und einer Krankheit, die mit dem Rauchen zusammenhängt - dem Herzinfarkt. Das Ziel dieser Studien ist der Nachweis einer Verbindung zwischen dem Rauchen, Biomarkern, die Zwischenstadien gestörter Stoffwechselwege abbilden, und dem Herzinfarkt. Der wissenschaftliche Beweis einer solchen Verbindung und dessen Art könnte unser Wissen über die Basis dieser Stoffwechselwege erweitern.

Die erste Studie, die in dieser Doktorarbeit behandelt wird, verfolgt das Ziel eines verbesserten Verständnisses des Einflusses von Rauchen und der Raucherentwöhnung auf das Metaboliten-Profil im Blutserum des Menschen. Während das Rauchen das Risiko vieler Krankheiten, wie auch des Herzinfarkts, steigert, senkt die Raucherentwöhnung das Herzinfarktrisiko in kürzester Zeit erheblich. Die in dieser Doktorarbeit dargestellten Ergebnisse zeigen signifikante Unterschiede der Metabolit-Profile zwischen Rauchern, ehemaligen Rauchern und Nichtrauchern. 19 der 21 Metaboliten, die durch den Vergleich von Rauchern und Nichtrauchern identifiziert wurden, waren reversibel. Diese Ergebnisse wurden darüber hinaus von einer prospektiven Studie, der KORA S4->F4, bestätigt. Die Umkehrbarkeit der

rauchabhängigen Auswirkungen auf die Genexpression und Metaboliten-Profile konnte durch die Anwendung der Netzwerkanalyse und der Einbindung rauchabhängiger Gene und Metaboliten übereinstimmend gezeigt werden. Die Reversibilität der durch das Rauchen hervorgerufenen Veränderungen der Metaboliten des Serums fällt auch mit einem verringerten Herzinfarktsrisiko zusammen. Dies eröffnet die Möglichkeit diese Metaboliten möglicherweise als Biomarker zu verwenden, um Raucherkrankheiten zu charakterisieren.

Inspiriert von der ersten Studie wurden zwei weitere Studien mit verschiedenen Zielen initiiert. Die zweite Studie der Doktorarbeit verfolgt das Ziel mehrstufige „-omics“ Daten zu verwenden, um den Einfluss von Rauchen auf die Metabolit-Profile, die mit DNA Methylierung und Genexpression zusammenhängen, darzustellen. Geeignete Biomarker von Rauchern wurden im Vorfeld unabhängig voneinander in Epigenomik, Transkriptomik und Metabolomik bestimmt. Anhand der Durchführung von Mediations-Analysen konnte die Möglichkeit einer Verbindung zwischen dem Rauchen und der DNA Methylierung, der Genexpression und den Metaboliten abgeschätzt werden. Im Allgemeinen zeigten 7 CpG Bereiche signifikante Mediationseinflüsse für die Expression des Genes *LRRN3*. Von diesen 7 zeigten 2 auch eine signifikante Assoziation mit den Metabolitkonzentrationen von LPC (18:2) und PC ae C34:3.

In der dritten Studie wurden drei Metaboliten (Arginin, LPC (17:0) und LPC (18:2)), die als neue Biomarker für das Auftreten eines Herzinfarkts dienen könnten, basierend auf einem gezielten Metabolomics-Ansatz in zwei prospektiven Studienkohorten identifiziert. Diese Metaboliten konnten signifikant in einem Cox Regressionsmodel nach der Anpassung der anderen Herzinfarktrisikofaktoren wie dem Rauchen und dem C-reaktiven Protein (CRP) mit dem Herzinfarkt in Verbindung gebracht werden.

Die Einbeziehung dieser Metaboliten in die etablierten Herzinfarktmodelle lieferte zusätzlich signifikant verbesserte Vorhersagen. Zusätzlich konnte die Beobachtung, dass diese Metaboliten mit CRP zusammenhingen, potenzielle Endzündungsprozesse, in denen diese involviert sind, anzeigen. Unter den drei oben genannten Metaboliten waren auch Arginin und LPC (18:2), die bereits in der ersten Studie der Doktorarbeit mit dem Rauchen in Verbindung gebracht werden konnten. Dies impliziert einen metabolischen Zusammenhang zwischen Rauchen und Herzinfarkt.

Zusammengefasst zeigt diese Doktorarbeit Metaboliten auf, die mit dem Rauchen und mit Herzinfarkt zusammenhängen. Unter Verwendung eines systembiologischen Ansatzes wurden die Auswirkungen von Rauchen auf die DNA Methylierung und Genexpression, welche dazugehörige Schwankungen der Metabolitkonzentrationen herbeiführen, durch die Einbindung von mehrstufige „-omics“ Daten analysiert. Die Metaboliten, welche mit dem Rauchen und mit Herzinfarkt assoziiert sind, könnten zu einem tieferen Einblick in die molekulare Basis der Verbindung zwischen Herzinfarkt und dessen Risikofaktor, dem Rauchen, beitragen.

Chapter 1 Introduction

After the first introduction of the field a decade ago, systems biology has greatly progressed and expanded in different aspects of the life sciences (Chuang *et al*, 2010). However, there is still not yet a concise and concrete definition of systems biology. While Ideker *et al.* emphasized a genome-wide systematic approach with mathematical modelling (Ideker *et al*, 2001), others suggested small scale quantitative studies (Tyson *et al*, 2001). Additionally, Hiroaki Kitano proposed the integration of experimental and computational approaches (Kitano, 2002). Despite the diversity of the definition, all opinions share a conceptually common aspect: the computational modeling of the interactions between different components of a biological system. This concept of the systems-level analysis is enabled by the high-throughput of 'omics' measurements, such as gene expression profiling, DNA methylation profiling and metabolite profiling, which provide comprehensive information of a biological system. In epidemiological studies, a systems biology approach using different 'omics' data can extend the view from observing epidemiological associations between risk factors and disease outcomes, for example smoking and MI, to the molecular level understanding of disease etiology. The studies in this thesis followed this concept to investigate smoking, cardiovascular diseases (CVDs), including MI, and their relationship with one another in the light of multi-level 'omics' data (Figure 1).

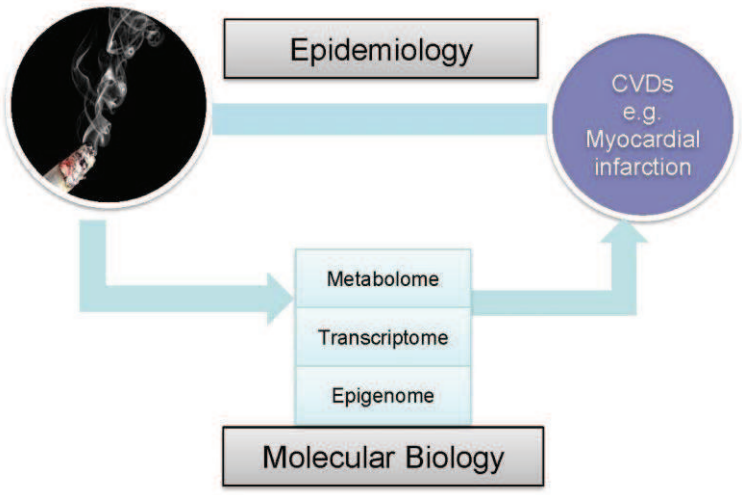

Figure 1. 'Omics' data reveals molecular basis for the links between smoking and CVDs

1.1 Linkage between smoking and CVDs

Smoking is a significant risk factor for CVDs as found in many epidemiology studies (Bøttcher & Falk, 1999; Howard *et al*, 1994; Libby, 2002; Szmitko *et al*, 2003). The benefits of smoking cessation are remarkable. The risks of CVDs are reduced in former smokers compared to current smokers (Danesh *et al*, 2000, 1999; Rigotti & Pasternak, 1996); the mortality and future cardiac events both decline in former smokers (Cook *et al*, 1986; Aberg *et al*, 1983). Nevertheless, for cancers, especially for adenocarcinoma, the risk remains high in former smokers compared to never smokers (Halpern *et al*, 1993; Ebbert *et al*, 2003). Studies have been made to find the molecular basis for the influence of smoking and smoking cessation on cardiovascular risks. It is known that smoking is associated with the increase of several CVDs related inflammatory biomarkers, e.g. CRP and fibrinogen (Bakhru & Erlinger, 2005; de Maat *et al*, 1996; Pradhan *et al*, 2002), while smoking cessation can largely reduce the level of these biomarkers (Yanbaeva *et al*, 2007). However, with this said, there is also evidence that other molecular changes associated with smoking are permanent which are related to a constant disease risk even after smoking cessation, for example,

loss of heterozygosity and hypermethylation in the promoter regions of cancer related genes (Powell *et al*, 1999; Wistuba *et al*, 1997, 2002; Guo *et al*, 2004; Beane *et al*, 2007).

1.2 Metabolomics in systems biology

As one of the emerging 'omics' of the post-genomic era, metabolomics focuses on the study of "*metabolites and low-molecular-weight intermediates, which are varying according to the physiology, developmental or pathological state of the cell, tissue, organ or organism*" (Functional genomics: lessons from yeast, 2002). Metabolomics is successfully applied in the study of complex phenotypes, such as obesity and aging (Barber *et al*, 2012; Wahl *et al*, 2013; Yu *et al*, 2012; Jourdan *et al*, 2012). It is also used to find the metabolites that can potentially be used therapeutically or diagnostically (Long *et al*, 2011; Wang-Sattler *et al*, 2012; Shah *et al*, 2012; Rhee & Gerszten, 2012; Patti *et al*, 2012; Floegel *et al*, 2012).

As with other 'omics' studies, metabolomics provides detailed information regarding the content of cells, tissues, organs or bio-fluids on a large-scale and high-throughput manner (Patti *et al*, 2012; Baker, 2011). Several different techniques exist for the high-throughput measurement of metabolites where each technique covers a certain panel of metabolites (Patti *et al*, 2012; Baker, 2011). Nuclear magnetic resonance (NMR) spectrometry and mass spectrometry (MS) are generally the two main analytical techniques which are used for the measurement of metabolite concentrations where different analytical approaches could be applied; the so-called targeted or non-targeted profiling approaches (Patti *et al*, 2012).

MS-based methods were used in the metabolomic measurements presented in this thesis; as such an analytical platform is available in Helmholtz Zentrum München. These MS-based methods include two major techniques: flow injection analysis MS

(FIA/MS) and high performance liquid chromatography MS (HPLC-MS) in combination with tandem mass spectrometry (MS/MS). By these techniques, around 160 to 180 metabolites can be measured. However, the scale and coverage of metabolomics is in no comparison to the other 'omics'. The estimated number of metabolites in humans remains under debate, ranging from thousands to tens of thousands (Baker, 2011). It is still impossible to measure the "*whole metabolome*" (Patti *et al*, 2012).

Metabolomics is an extension of other 'omics', such as transcriptomics and proteomics, which are used to elucidate the functions of genes. One of the main goals for the researchers in this field is to find the associations between metabolites and genetic variations (Gieger *et al*, 2008; Illig *et al*, 2009; Suhre *et al*, 2011), ultimately aiming to understand the differences in the genetics associated with inter-individual differences in metabolism. Other researchers focused on the associations of metabolomic traits with epigenomic traits (Petersen *et al*, 2014) and gene expressions (Knolhoff *et al*, 2013; Wilmes *et al*, 2013), which provide further insights into the regulation of human metabolism. Metabolite profiles are successfully integrated with other 'omics' data to model regulation pathways (Zhu *et al*, 2012; Cavill *et al*, 2011). Together with other 'omics', metabolomics provides a powerful tool to analyze physiological and disease-induced biological states at the molecular level, taking into account both the organism's intrinsic properties, i.e. genetic factors and environmental influences.

1.3 Metabolomic study of smoking and CVDs

The metabolomic approach provides a functional read-out of activities located downstream of the proteome that are more closely related to the physiological status, and thus may be particularly useful for the study of both environmental influences

(Ellis *et al*, 2012) and diseases outcome (Wang *et al*, 2011a). It is of the believe that studying smoking, which is a strong risk factor for CVDs as well as for MI, will be a powerful approach for understanding the linkage between environmental exposure, metabolite profiles and disease outcome.

In human lung epithelial cells, differences of metabolite concentrations were found in various pathways when current smokers with never smokers were compared to one another, e.g. the urea cycle and lipid metabolism exposed to smoke (Vulimiri *et al*, 2009). In a pilot study with 283 male participants from the Cooperative Research in the Region of Augsburg (KORA) F3 study in Germany, it has been shown that levels of diacyl-phosphatidylcholines (diacyl-PCs), except for acyl-alkyl-PCs, were higher in 28 current smokers compared to that of a 101 never smokers (Wang-Sattler *et al*, 2008). The lower ratios of acyl-alkyl- to diacyl-PCs in current smokers might be regulated by the enzyme alkyl-dihydroxyacetone phosphate (DHAP) in both ether lipid and glycerophospholipid pathways (Wang-Sattler *et al*, 2008). However, little has been reported about the reversibility of the metabolite profile upon smoking cessation, which is important for the comprehensive understanding of the effects of smoking (Xu *et al*, 2013). It is also known that the metabolite profiles are different between men and women (Mittelstrass *et al*, 2011), thus the metabolic response to smoking may also be gender specific.

Metabolomics has been used as a novel tool to study CVDs (Shah *et al*, 2010; Magnusson *et al*, 2013). Variations in metabolite profiles are associated with risk factors of CVDs, such as diabetes (Wang-Sattler *et al*, 2012; Wang *et al*, 2011a) and smoking (Wang-Sattler *et al*, 2008; Xu *et al*, 2013). Recently, it was found that the gut flora metabolism of lipids will promote CVDs, shedding light on a novel

pathomechanism which leads to atherosclerosis (Wang *et al*, 2011b). Another study has reported branched-chain amino acids and urea cycle metabolites as independent cardiovascular risk factors (Shah *et al*, 2010). Urinary excretion of kynurenine and tryptophan were found to be associated with incident major coronary events and acute MI (Pedersen *et al*, 2013). All of these findings emphasized the capacity of metabolomics as a tool to explore disease mechanism.

1.4 Systems biology study of smoking and CVDs

A large number of studies have been conducted to find the influence of smoking at genome, transcriptome and epigenome. Genome wide meta-analysis of over 70K participants found loci which are related with smoking behavior, including loci which are also associated with initiation of smoking and cigarette smoked per day (Consortium, 2010). Genome-wide DNA methylation studies identified CpG sites which are related to smoking effects on functions of the immune, cardiovascular, tumorigenic or reproduction system (Zeilinger *et al*, 2013). At the transcriptomic level, one study found 175 smoking-related genes, by analyzing gene expression profiles in large airway epithelial cells. These genes were further classified into rapidly reversible, slowly reversible and irreversible genes (Beane *et al*, 2007). These discoveries provide the basis to model smoking behavior and to unveil the mechanism of the influence that smoking has on metabolite profiles using systems biology approaches. One study used an ontology-based computational representation to integrate prior knowledge in large-scale genetic association studies of smoking addiction (Thomas *et al*, 2009). Smoking associated SNPs were used in Mendelian randomizations studies to find causal roles of smoking in depression and body mass index (BMI) (Munafò & Araya, 2010; Freathy *et al*, 2011). However, the effects of

smoking regarding the initiation and propagation across different 'omics' networks are still not fully understood.

For CVDs, the genome-wide association studies (GWASs) have identified around a hundred loci for coronary artery diseases (Deloukas *et al*, 2012), which is the most common type of CVDs and leads to MI as well as heart failure. Large epigenomic and transcriptomic studies provided new gene expression and methylation signatures for CVDs (Schnabel *et al*, 2012) . Systems biology approaches have also been applied in the study of CVDs, for example a network based approach was used to identify functional modules of genes in CVDs (Lusis & Weiss, 2010). Neither CRP nor high density lipoprotein (HDL) cholesterol, by means of Mendelian randomization studies, showed significant causal associations with CVDs (Smith & Ebrahim, 2003). However, there is a causal association between low density lipoprotein (LDL) cholesterol with CVDs (Voight *et al*, 2012). Even though a lot of researches have been done, much still needs to be unveiled for the holistic picture of the underlining mechanism at play (Nabel & Braunwald, 2012).

1.5 Application of systems biology methods in epidemiology

1.5.1 Study design

Different study designs provide information in a different manner with regards to quality and quantity, which essentially influences the analysis and interpretation of the results (Pearce, 2012). To address the questions arising in epidemiology such as disease prediction or the influence of smoking, a proper design of study is crucial. In general there are two types of study designs which basically depend on whether interventions were used or not: observational and experimental. In this thesis only the observational study design is used.

The most common epidemiological study designs are the cross-sectional study, case-control study and cohort study. A cross-sectional study collects samples of a population at a defined time. It is less expensive and relatively easy to conduct (Pearce, 2012). But, the information it provides is only observational, thus unable to address questions such as causation or prediction. In a case-control study, people with disease or phenotype of interest are recruited as cases, and people who do not have a disease of interest or the specific phenotype, but with similar characteristics in other phenotypes will be recruited as controls (Pearce, 2012). The aim of such a design is to find the cause of diseases, especially rare diseases. However, it is usually done retrospectively, thus suffers from the problem of recall errors. A cohort study is a type of longitudinal study, which extends a survey over time. It can be conducted prospectively or retrospectively based on historical records. In a prospective cohort, researchers study a group of people without diseases or phenotypes of interest at the beginning of the study and collects information of the disease risk factors. After a certain period of time, disease information of the studied participants will be collected. This allows researchers to establish the time sequence of events, which can strongly aid with the study of causal associations. In comparison to the retrospective cohort, prospective cohort can reduce the recall error. The results provided by a prospective cohort are considered to be the most reliable in observational studies (Pearce, 2012).

However, the prospective cohort studies are expensive to conduct. To overcome this disadvantage some variations of the cohort study were developed, for example the case-cohort study (Prentice, 1986). In a case-cohort study design, a sub-cohort is randomly selected from the parent cohort. At the end of the study, all cases in the parent cohort are also added into the sub-cohort which forms the sample for analysis (Prentice, 1986). Owing to the fact that less samples are required in this design it is

considered to be less expensive, more efficient and flexible to conduct than a cohort study whilst having similar powerful outputs (Barlow *et al*, 1999)

In this thesis the cross-sectional study, case-control study, prospective cohort study and case-cohort study design were used.

1.5.2 Statistical modeling and data analysis

Besides describing the statistics of the study population, most questions in epidemiological studies, such as identifying the risk factors of health related problems, evaluating therapy, etc., can be modeled as finding the associations between risk factors and diseases outcome. Regression models and correlations are the most commonly used methods in practice to unveil such associations (Rothman, 2012).

However, a simple correlation between two factors or a linear regression of the outcome on the risk factors is in most cases not sufficient to conclude a direct association between two factors. In observational studies, a most common problem is confounding. A significant association between two factors in a statistical model can be spurious when both factors are not associated directly to one another but commonly caused by a third variable, naming a confounding factor. In epidemiology, the inability to control interpersonal variations of participants, makes confounding a particular challenge. It is common practice to add the confounding factors in the regression models to control for the effect of a particular risk factor, such as smoking.

The relationship between diseases and risk factors in epidemiological research are not a simple one-to-one correlation. It should rather be illustrated in a form of a complex network with interaction between different genes, proteins, metabolites and other molecules (Barabási *et al*, 2011; Vidal *et al*, 2011; Chan & Loscalzo, 2012). To understand the biological processes for the pathomechanism of a disease, the

relationship gets even more complicated as a single level network is usually not enough to depict the whole picture of disease pathways. Integration of the multi-level 'omics' data, which is also an issue of methodology development in systems biology, is needed to provide a systemic view of diseases (Barabási *et al*, 2011; Moreau & Tranchevent, 2012). One method to infer such a network of linear dependencies among a set of variables, is to compute all pairwise Pearson correlations or Spearman correlations (Langfelder & Horvath, 2008; Carter *et al*, 2004). More sophisticated methods can illustrate complicated relationships, when the confounding factors are also considered as mentioned previously. Davey-Smith, *et.al.*, conducted a series of studies using Mendelian randomization analysis to find the causal associations between risk factors and diseases (Smith & Ebrahim, 2003; C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC), 2011; Davey Smith, 2011). The method of Mendelian randomization analysis enables causal inference in observational studies, which for a long time could only be addressed by randomized controlled trails (Smith & Ebrahim, 2003). These findings pinpointed the causal roles of risk factors in disease development, assisting in the process to find a proper biomarker for intervention studies. Some other studies used mediation analysis to quantify the influence of one intermediate phenotype on the association between one exposure or genetic variation and the disease; such an example is the effect of ABO blood type and CVDs (Chen *et al*, 2014). It can also make causal inference of disease regulatory networks, integrating genomic, transcriptomic and epigenomic data (Bellavia *et al*, 2013; Wardle *et al*, 2008). Other methods, such as Bayesian networks, were also applied to find interactions between different 'omics' levels in systems biology studies (Zhu *et al*, 2012; Sebastiani *et al*, 2007).

1.5.3 Databases and bioinformatics tools

To perform the modeling and data analysis, a set of tools are available.

R is a comprehensive platform for statistical computation and graphics (R Core Team, 2013). It was derived from the statistical computing platform S (Becker & Chambers, 1984) and Scheme (Clinger & Rees, 1991). It has now a large user and developer community, providing a variety of packages for different statistical modeling purposes, such as the package "survival" for survival analysis, "lme4" for linear mixed effect models, "sem" for structural equation models and for Mendelian randomization studies.

For metabolomics analysis, several web-based analytic tools were developed. The metP-server is a tool developed by a group from the Helmholtz Zentrum München, providing automated and standardized data analysis for quantitative metabolomics data, covering quality controls, hypothesis testing, correlation analysis, PCA etc. (Kastenmüller *et al*, 2011). It provides special support to analyze the original measurement data from Biocrates Absolute*IDQ* kits. A more comprehensive platform is the MetaboAnalyst (Xia *et al*, 2009). It provides a tool suite to facilitate metabolomics data processing and statistical analysis. In addition to the functions provided by metaP-server, it supports peak detections from original spectra from MS and NMR measurements; provides additional tools for statistical analysis, including machine learning methods such as SVM and random forest; and gives better support with regards to functional analysis by providing enrichment analysis and metabolic pathway analysis (Xia *et al*, 2009).

Several databases are available for functional analysis of metabolomics studies. The Human Metabolome Database (HMDB) is currently the most comprehensive curated collection of human metabolite and human metabolism data (Wishart *et al*, 2009). It

contains 2180 endogenous metabolites, with structural information, disease associations, enzyme data, and pathway information etc., derived from literature and experimental metabolite concentration data (Wishart *et al*, 2009). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for the understanding of high-level functions of biological systems (Kanehisa *et al*, 2014; Kanehisa & Goto, 2000). It provides an a systemic view of biological processes and pathways with multi-level 'omics', consisting of genes, proteins, metabolites and other chemical substances that are integrated by wiring diagrams of interaction, reaction and relation networks. The database includes only data with solid evidence, which are manually curated, thus covering only a limited portion of pathways and a glimpse of the whole human metabolism (Kanehisa *et al*, 2014; Kanehisa & Goto, 2000). Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (Szklarczyk *et al*, 2011) includes a database of protein-protein interactions (PPIs) from various sources and is essentially a tool to search for PPIs for a given set of proteins. As an extension of STRING, Search Tool for Interactions of Chemicals (STITCH) (Kuhn *et al*, 2012) connected over 300000 chemicals to the PPI network from 1133 organisms. Both tools provide a systematic view of the interactions between genes, proteins, and metabolites, linking metabolomics information with other 'omics'. These tools have been successfully applied in the study of drug-target interactions (Meslamani & Rognan, 2011; Kalinina *et al*, 2011) and also metabolomics studies (Wang-Sattler *et al*, 2012).

Chapter 2 Material and methods

2.1 Population based KORA cohort study

The KORA surveys are population-based studies conducted in the region of Augsburg, Germany (Holle *et al*, 2005; Wichmann *et al*, 2005). Four surveys were conducted with 18079 participants recruited from 1984 to 2001. Four KORA studies (S4, F4, F3 and S2) were used in the thesis.

The KORA survey 4 (S4) is a population based cohort of 4261 individuals (aged 25-74 years) examined from 1999 to 2001. From 2006 to 2008, 3080 participants (with an age range of 32-81) took part in a follow-up (F4) survey. The KORA survey 2 (S2) consist of 4940 participants enrolled between 1989 and 1990. CVD events for the participants were identified up to 2002. A case-cohort was established which included a randomly selected sub-cohort (n = 813) and all incident cases of MI during the follow-up study (n = 146). The KORA F3 was conducted between 2004 and 2005 as the follow up study of the KORA survey 3 (S3), which was conducted between 1994 and 1995. Samples from 821 participants were used in this thesis. The ethics committee of the Bavarian Medical Association in Munich, Germany approved the KORA studies. Each participant in these studies was given a written consent form and completed a lifestyle questionnaire requesting information on a number of parameters such as smoking status (current, former, never), medical history, alcohol consumption, physical activity, etc. Additionally a standardized medical examination, including blood pressure measurements and anthropometric measurements, was performed on all the participants (Holle *et al*, 2005). During the examination, serum samples were collected in parallel during the examination (described in detail in the next section).

2.2 Blood sample collection

To measure the metabolite concentrations in human blood, serum samples were collected from the KORA participants. The blood was drawn into SMonovettes tubes (SARSTEDTAG & Co., Nümbrecht, Germany) in the morning between 08:00 and 10:30 after a period of overnight fasting for at least eight hours. Serum tubes were gently inverted twice, followed by 30 min resting period at room temperature to ensure complete coagulation. The tubes were then centrifuged at 2750 g at 15°C for 10 min. Plasma and serum were filled into synthetic straws, which were stored in liquid nitrogen until the metabolic analyses were conducted. Serum samples from the KORA S2, F3, S4 and F4 were used in the analysis (Wang-Sattler *et al*, 2012; Mittelstrass *et al*, 2011; Jourdan *et al*, 2012).

In the first study presented in this thesis, serum samples from the KORA S4 were used to illustrate the effect of smoking on the metabolite profiles in a cross-sectional manner whilst the follow-up study of KORA S4 (KORA F4) was used to verify the results in a longitudinal setting. In the second study, in addition to the metabolomic data, gene expression and DNA methylation profiling data from KORA F4 were used to analyze the effects of smoking on multi-level 'omics', while KORA F3 was used as the replication cohort. In the third study in this thesis, KORA S4, with MI registry information until 2009, was used as the discovery dataset whilst the KORA S2 case-cohort served as the replication study.

2.3 Quantification of metabolite profiles

2.3.1 Absolute*IDQ* kit p150/p180

The metabolite profiling was performed using the Biocrates Absolute*IDQ*^{TM} kit p150 and p180. The Absolute*IDQ*^{TM} kit p150 used FIA-MS/MS techniques. This technique has been described in detail elsewhere (Weinberger & Graber, 2005; Weinberger, 2008). Briefly, the assay preparation was done by an automated robotics system

(Hamilton Robotics GmbH) on special double-filter plates with 96 wells. These plates also contain the isotope labeled non-radioactive internal standards, blank samples (PBS) and quality controls. Assays used 10 µl serum or plasma samples together with phenylisothiocyanate (PITC)-derivatisation of amino acids, which was further extracted by means of an organic solvent and several other liquid handling steps. FIA MS/MS on an API 4000 QTrap instrument (Applied Biosystems) was used for the quantification of amino acids, acylcarnitines, sphingomyelins, phosphatidylcholines, and hexose. Concentrations were calculated and evaluated in the MetIQ software provided by the manufacturer. It compared measured analytes in a defined extracted ion count section to those of specific labeled internal standards or non-labeled, non-physiological standards (semi-quantitative) provided by the kit plate. This method has been proven to be in conformance with the "Guidance for Industry — Bioanalytical Method Validation" published by the FDA (Food and Drug Administration), which allows for a given error range with regards to the reproducibility of the analysis (Altmaier *et al*, 2011; Römisch-Margl *et al*, 2012).Serum samples from the KORA F3 and from the KORA F4 were measured using this particular kit for metabolite concentration profiles.

The Absolute*IDQ*™ kit *p*180 is an upgrade of the Absolute*IDQ*™ kit *p*150. It uses the combination of FIA-MS and HPLC-MS to detect metabolite concentrations. Metabolite concentrations were measured using the Absolute*IDQ*™ kit *p*180 according to the manufacturer's instructions on an API4000™ LC-MS/MS system which is equipped with an electrospray ionization source. Samples (10 µl) were pipetted onto the spots of the kit plate. The plate was centrifuged at 100 g for 2 min, receiving about 250 µl sample in plate 1 (FIA plate). The upper plate was removed, and 150 µl of each sample was transferred into a second plate (LC-MS plate). HPLC-

grade water (150 µl) was added to the LC-MS plate, and 500 µl of MS running solvent (Biocrates solvent diluted in methanol) was added to the FIA plate. The LC-MS plate was measured first by scheduled multiple reaction monitoring. The FIA plate was stored at 4°C. Concentrations were calculated and evaluated in the Analyst/MetIQ software by comparing measured analytes in a defined extracted ion count section to those of specific labeled internal standards or non-labeled as well as non-physiological standards (semi-quantitative) provided by the kit plate (Schmerler *et al*, 2012). The serum samples from the KORA S2, S4 were measured using this kit for metabolite concentration profiles.

2.3.2 Metabolite panel

In total, up to 190 different metabolites were quantified by these two kits. Absolute*IDQ*™ kit *p*150 measured 163 metabolites, including 14 amino acids (13 proteinogenic and ornithine), total hexose (around 90 – 95% glucose), free carnitine (C0) and 40 other acylcarnitines (Cx:y), 15 sphingomyelins (SMx:y), 77 phosphatidylcholines (PCs, diacyl (aa) and acyl-alkyl (ae)) as well as 15 lysophosphatidylcholines (LPCs). The lipid side chain composition is abbreviated as Cx:y, where x denotes the number of carbons in the side chain and y denotes the number of double-bonds. The Absolute*IDQ*™ kit *p*180 measured 186 metabolites, including 21 amino acids (19 proteinogenic, citrulline and ornithine), hexose, free carnitine, 39 acylcarnitines, 15 sphingomyelins, 90 phosphatidylcholines (14 LPCs and 76 PCs) as well as 19 biogenic amines. There were 159 overlapping metabolites from the two kits.

2.3.3 Quality controls for metabolite measurement

Different quality control processes were introduced for metabolite profiles measured by different Biocrates Absolute*IDQ*™ kits. The metabolite data quality control

procedure for the KORA S4 samples was described in our recently published work (Wang-Sattler *et al*, 2012). 140 metabolites passed the two quality controls: one hexose (H1), 21 amino acids, eight biogenic amines, 21 acylcarnitines, 13 sphingomyelins (SMs), eight lyso-PCs and 33 diacyl-PCs (PC aa Cx:y), and 35 acyl-alkyl-PCs (PC ae Cx:y). As mentioned before, the lipid side chain composition is abbreviated as Cx:y. Concentrations of all analyzed metabolites are reported in μmol/L (μM). The data cleaning procedure for the KORA F4 samples was described in detail in literature (Mittelstrass *et al*, 2011; Yu *et al*, 2012). 123 metabolites passed the quality control criteria in both S4 and F4, and were used for the prospective study.

For the metabolite measurement in KORA S2, the same quality control criterion as for KORA S4 was applied. This resulted in 134 metabolites passing the criteria: one hexose (H1), 19 acylcarnitines, 20 amino acids, four biogenic amines, 13 SMs, 34 diacyl- phosphatidylcholines (PC aa Cx:y), 35 acyl-alkyl-phosphatidylcholines (PC ae Cx:y) and eight lysophosphatidylcholines (LPC (x:y)).

The metabolites used in KORA S4, F4 and S2 study are listed in the Table S1 in the Appendix.

2.4 Gene expression profiling

Peripheral blood, for gene expression profiling, was drawn under fasting conditions from 599 KORA S4 individuals during the same time the serum samples, used for metabolic profiling, were prepared. Blood samples were collected directly in PAXgene (TM) Blood RNA tubes (PreAnalytiX). The RNA extraction was performed using the PAXgene Blood miRNA kit (PreAnalytiX). Purity and integrity of RNA was assessed on the Bio-analyzer (Agilent) with the 6000 Nano LabChip reagent set (Agilent). In all, 500 ng of RNA was reverse transcribed into cRNA and biotin-UTP labeled, using the Illumina TotalPrep-96 RNA Amplification Kit (Ambion). From this

3000 ng of cRNA was hybridized to the Illumina HumanHT-12 v3 Expression BeadChip. Chips were washed, detected and scanned according to manufacturer's instructions. Raw data were exported from the Illumina 'GenomeStudio' Software to R. The data were converted into logarithmic scores and normalized using the quantile method (Bolstad *et al*, 2003; Wang-Sattler *et al*, 2012).

2.5 Methylation profiling

DNA Methylation profiling was performed as described in a previous publication (Zeilinger *et al*, 2013). In total, methylation profiles for 1802 samples from the KORA F4 and 500 samples from the KORA F3 were measured. Three samples which had less than 80% high quality probes (detection p-value <0.01) were excluded. CpG sites which were in close proximity (50 base pairs (bp)) to common SNPs were also removed. Color bias adjustment based on a smooth quantile normalization method as well as background level correction based on negative-control probes were performed for each chip using the R "lumi" package (Du *et al*, 2008). Data were then normalized following the BMIQ pipeline (Teschendorff *et al*, 2013). The M-value of each methylation probe was used for data analysis, which is calculated as the log2 ratio of the intensities of methylated probe vs. unmethylated probe.

2.6 Statistical analysis

All statistical analyses were performed using R statistical environment (R Core Team, 2013).

2.6.1 Correlations

A correlation exists between two variables when one of them is related to the other. Pearson's (product moment) correlation coefficient (r) measures the strength of the linear relationship between the paired x- and y-quantitative values in a sample if both x and y follows a normal distribution (Triola *et al.*, 2006). For scenarios where the distribution is not normal, Spearman's correlation (ρ) can be calculated instead.

2.6.2 Generalized linear model

In this thesis, the generalized linear model was used to assess the associations between outcome and risk factors. This model is used to assess the relationships between metabolite concentrations and smoking exposure, and also the associations between CRP level and metabolite concentrations. The outcome of an observational study, either diseases or intermediate phenotype, can be considered as a result of a linear combination of the effects of the independent variables, environmental exposure or disease risk factors. This outcome can be described in the form of a linear model,

$$\boldsymbol{Y} \sim g^{-1}(\boldsymbol{X\beta}) + \varepsilon$$

, where X is the independent variable; ε indicates the error term whilst g is the link function, which describes the relation between the linear predictor $\boldsymbol{X\beta}$ and the outcome $\boldsymbol{Y}$. Depending on the different types of outcomes and the relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$, g can be designed differently (Woodward, 1999).

Two types of link functions were applied for the studies in this thesis. For outcome follows a normal distribution, the linear regression model is used, which is the simplest form of the generalized linear model with a link function as following.

$$g(\boldsymbol{X\beta}) = E(Y)$$

This model can be used to analyze the association between $\boldsymbol{Y}$ and $\boldsymbol{X}$ which follows a linear-response relation. For binomial outcome data, where $\boldsymbol{Y}$ indicates yes/no for disease or exposure, the logit function was used.

$$g(\boldsymbol{X\beta}) = \ln\left(\frac{E(Y)}{1 - E(Y)}\right)$$

This is also known as the logistic regression model (Woodward, 1999).

Several different approaches are available to estimate the parameters $\boldsymbol{\beta}$ for the models. While the outcomes of the samples are uncorrelated, for example, in a cross-sectional study where the samples are independently drawn from a population, ordinary least squares (OLS) can be used for parameter estimation. This method minimizes the sum of squared distance between the observed outcome in the dataset and the predicted outcome by the linear estimation. Under the same assumption, for logistic regression, the maximum likelihood estimation (MLE) is used for parameter estimation (Woodward, 1999).

However, in longitudinal studies, outcome variable for one participant are measured repeatedly at multiple time points. The observations between two consecutive time points are usually correlated, thus estimation methods need to take these correlations into account. Liang, *et.al* proposed the Generalized Estimation Equation (GEE), which is an extension for GLM in such a situation. They proved in their paper that parameter estimation using OLS or MLE for data with a correlated outcome is less efficient compared to GEE (Liang & Zeger, 1986). In the case of linear regression, instead of estimating the coefficient estimators as

$$\boldsymbol{\beta} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$$

, GEE incorporates the correlation matrix of observations from the same participants $\boldsymbol{W}$ in the estimation equation as,

$$\boldsymbol{\beta}' = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{Y}$$

2.6.3 Linear mixed effect model

The linear mixed effect model is another method to assess the association between outcome and risk factors in longitudinal studies. In this thesis, to assess the role of smoking cessation for the smoking quitters, who were current smokers at S4 but former smokers at F4, the linear mixed effect model was used to account for the effect of repeated measurements. The model contained the fixed effect of smoking status (current smokers, former smokers and never smokers), age, BMI, and alcohol consumption with a random effect assigned to each participant.

2.6.4 Cox proportional hazard model

Cox regression models were used to assess the association between metabolite concentration and incident MI. Initial analyses used a crude model adjusting for age and sex, followed by a multivariate model adjusting for body mass index (BMI), smoking status, alcohol consumption, diabetes, systolic blood pressure, high density lipoprotein cholesterol (HDL-C), and total cholesterol. In the final full model CRP was included to control for inflammatory status. Significance level was set at $P<0.05$. In the KORA S2 case-cohort a sex stratified weighting was used in the Cox regression models to account for the over-sampling of the case-cohort design (Barlow *et al*, 1999). The significance level was set at a p-value <0.05. Sensitivity analysis was conducted to analyze the effects of statin medication on diabetes.

2.6.5 Model evaluations

R^2 and AIC was used to evaluate the goodness of fit for each model (Akaike, 1974). To compare two nested models for the goodness of fit, the likelihood ratio test was used (Huelsenbeck & Crandall, 1997).

Receiver operating characteristic (ROC) curve was used to compare the prediction performance with respect to MI of the different adjustment models in the study.

Added predictive values of potential biomarkers for incident MI were indicated by the increase in the C statistics, net reclassification index (NRI) and integrated discrimination improvement (IDI) statistics (D'Agostino *et al*, 2008), which were assessed in all models and compared with the Framingham score.

2.6.6 Variable selection and model optimizations

Potential biomarkers for MI were identified via a two-step process from significant ($P<0.05$) associations in the full model. Significant metabolites were included in a lasso (L1)-regularized estimation. Finally from the non-zero terms in the first step potential biomarkers were identified using backward stepwise regression with Akaike's information Criterion (AIC).

2.6.7 Network analysis

Protein-protein interactions from the database of STRING (Szklarczyk *et al*, 2011) and enzyme-metabolite relations from the database of HMDB (Wishart *et al*, 2009) were retrieved to construct protein-metabolite networks. The protein-metabolite network contains links between metabolites, enzymes and coronary artery disease-related genes. Genes and metabolites were connected to one another by allowing at most one intermediate enzyme with Dijkstra's algorithm (Dijkstra, 1959), and further optimized by eliminating edges with STITCH scores less than 0.7. Each edge in the networks was manually checked. This methodology have been successfully implemented in our previous studies (Wang-Sattler *et al*, 2012; Xu *et al*, 2013). The analysis was performed using the R package igraph (Csardi & Nepusz, 2006) and the network was visualized using Cytoscape 2.8 (Smoot *et al*, 2011).

A similar approach was implemented in the network analysis of smoking related metabolites and smoking gene expression., Pathway enrichment analysis were additionally performed using MetaboAnalyst (Xia *et al*, 2012).

2.6.8 Mediation analysis

The associations between smoking, methylation, gene expression and metabolite were assessed using mediation analysis to find DNA methylations sites and gene expressions that mediated the effects of smoking on metabolite concentrations. Mediation analysis has already been used in causal inference in several studies (Liu *et al*, 2013; Schadt *et al*, 2005) to disentangle the complex relationship between different 'omics' data.

Mediation analysis can briefly be illustrated as follows:

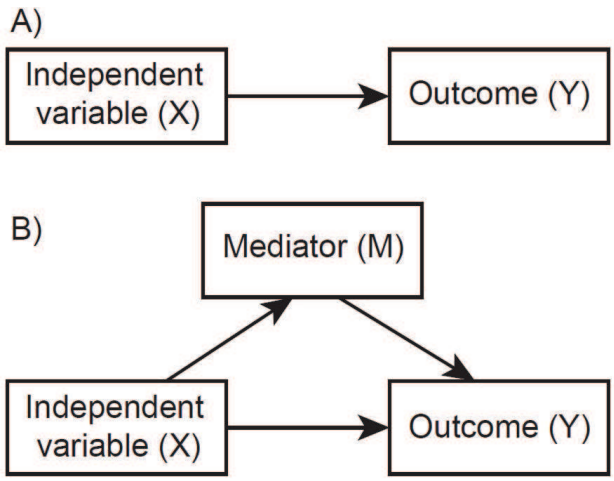

Figure 2. Mediation analysis

A) the association between outcome Y and independent variable X; B) the association between outcome Y and X when the association is mediated by M.

If we find significant association between the outcome *Y* and the independent *X,* as shown in Figure 1 A):

$Y \sim \alpha X$

, we can further conduct regression analysis to find if a third factor (*M*) mediates the association between *Y* and *X* by,

1) Regression of the mediator (*M*) on *X*

$M \sim \beta X$

2) Regression of *Y* on both *X* and *M*

$Y \sim \alpha' X + \gamma M$

If the association between *M* and *X*, and the association between *Y* and *M* are both significant, while the effect size α is smaller than α', we can suggest M mediated the association between X and Y. The effect size of the mediation is calculated as,

$Mediation\ effect = \alpha - \alpha' = \beta\gamma$

The meditation effect can be transformed into a t-statistics and test for significance,

$T = (\alpha - \alpha')/SE$

$SE = \sqrt{(\beta^2\sigma_\gamma^2 + \gamma^2\sigma_2^\beta)}$

, which is known as the Sobel test (Sobel, 1982).

The mediation can account for all associations between the outcome and independent variables (after adding M in the model), if the association between X and Y are not significant anymore and the associations between X and Y are close to zero, after adding M in the model.

Chapter 3 Results

This chapter shows the results of the three studies in this thesis and is divided into three parts. The first part presents the results from the study of effects of smoking and smoking cessation on human metabolite profile. The second study was extended to multi-level 'omics', integrating transcriptomic, epigenomic, and metabolomic data to find interactions between DNA methylation, gene expression and metabolite, which mediate the effects of smoking on metabolite concentrations. The last section presents a study using metabolomics approach to find potential metabolite biomarkers for incident MI as well as to understand the underlying disease mechanism.

3.1 Effects of smoking and smoking cessation on metabolite profile

In this study, the associations between smoking and the concentrations of metabolites in 1241 serum samples from the KORA baseline S4 and follow-up F4 study were investigated, aiming 1) to extend the knowledge of smoking associated metabolites beyond our pilot study (Wang-Sattler *et al*, 2008) by including female current smokers at two time points over seven years, 2) to investigate whether smoking-related changes in metabolite profile are reversible after smoking cessation, and 3) to provide insights into the pathophysiological consequences of smoking in protein-metabolite networks.

3.1.1 Characteristics of participants of the cross-sectional KORA S4

In the KORA S4 study, serum samples from 1614 persons aged between 55 to 74 years were available. Furthermore metabolite concentrations of serum samples from 1036 participants were measured in both KORA S4 and F4. Participants with non-fasting status (N = 216) or missing values (N = 22) were excluded from the analysis; 145 persons in the KORA S4 and 116 persons in the KORA S4 → F4 were further excluded, whose spouses were current smokers, to rule out passive smoking effects.

Participants were divided into three groups, current smokers, former smokers and never smokers according to their self-reported smoking status. Population characteristics were shown in Table 1. On average, current smokers were two to three years younger and had a lower BMI than former smokers and never smokers. Male current smokers showed higher alcohol consumption (27.5 g/day) than male never smokers (20.5 g/day), but there were no significant difference observed in women. Furthermore, the statistics showed differences in lifestyle factors between men and women. Alcohol consumption was higher in men than women (p-value = 1.5E-11 (current smokers); p-value = 2.2E-18 (former smokers); p-value = 9.5E-17 (never smokers)), and the smoking intensity (in pack years) was higher in male than in female current smokers (p-value = 6.0e-6).

Table 1. Characteristics of the cross-sectional KORA S4.
The study characteristics of KORA S4 are shown separately for current smokers, former smokers and never smokers. Values are shown as Mean ± standard deviation (SD) when appropriate. †P-values are calculated by student's t-test.
Explanation of variables: BMI, body mass index, in kg/m^2; alcohol, gram of alcohol consumption per day, in g/day; pack year, calculated by: the number of cigarettes consumed per day × years of smoking / 20; quit time, the time till the survey is conducted since the person has stopped smoking, in years; CS: Current Smoker; FS: Former Smoker; NS: never smoker

	CS	FS	NS	P-value†	
				CS vs FS	CS vs NS
Male (N=646)					
N (%)	125 (19.3%)	321 (49.7%)	200 (31.0%)		
Age (years)	62.2 ±5.3	65.3 ± 5.3	64.1 ± 5.6	7.9E-08	3.0E-03
BMI (kg/m²)	27.0 ± 3.6	28.9 ± 3.6	27.8 ± 3.4	1.5E-06	6.5E-02
Alcohol (g/day)	27.5 ± 29.0	24.1 ± 24.3	20.5 ± 21.3	0.25	0.02
Pack years	39.3 ± 22.4				
Quit time (years)		23.6 ± 12.6			
Female (N=595)					
N (%)	70 (11.8%)	130 (21.8%)	395 (66.4%)		
Age (years)	61.3 ± 5.2	64.0 ± 5.2	64.6 ± 5.3	7.5E-04	5.9E-06
BMI (kg/m²)	27.2 ± 4.5	28.7 ± 5.0	28.5 ±4.6	0.029	0.02
Alcohol (g/day)	6.5 ± 10.9	10.0 ±12.8	7.5 ± 11.1	0.042	0.48
Pack years	25.8 ± 15.3				
Quit time (years)		20.9 ± 13.1			

3.1.2 Metabolomic differences between current, former and never smokers

The effects of smoking were assessed in the cross-sectional KORA S4 study and verified in the longitudinal KORA S4→F4 (Figure 3).

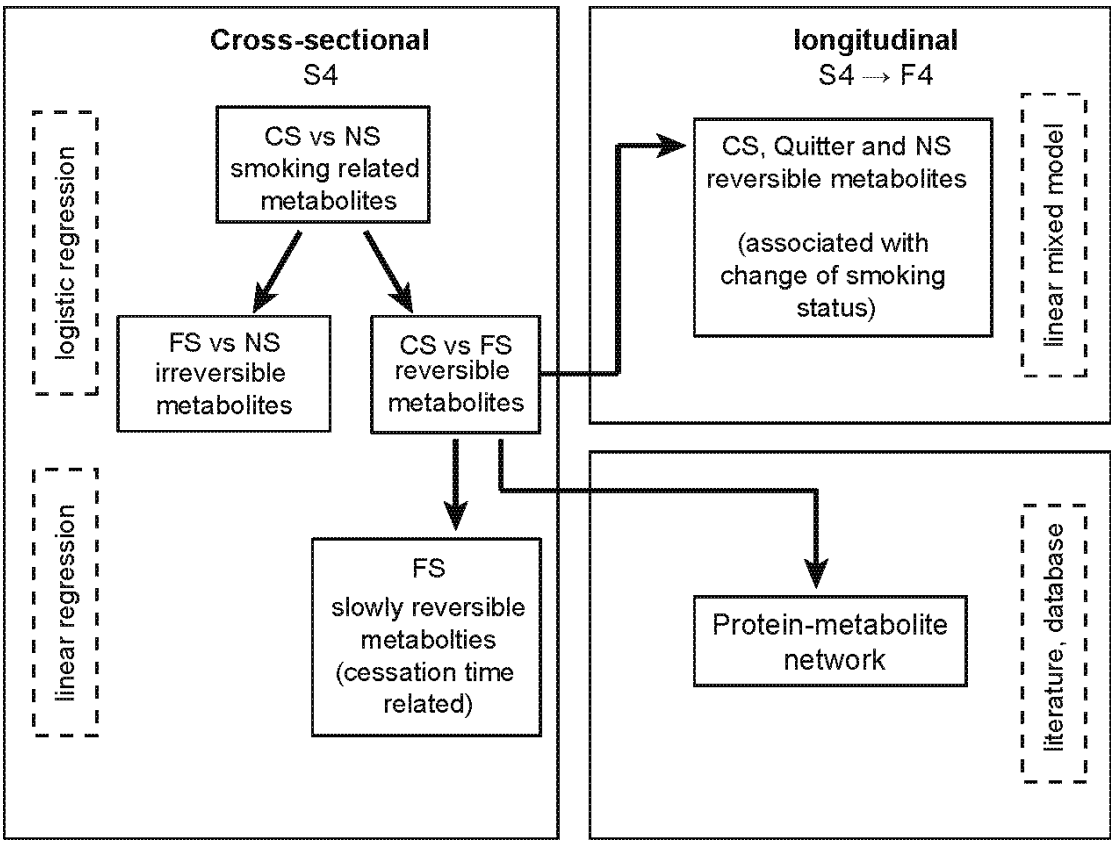

Figure 3. Flow diagram illustrating the analysis strategy for effects of smoking on metabolite profile

Abbreviations: CS: current smokers; FS: former smokers; NS: never smokers.

The analysis showed 18 metabolites in men and six in women that were significantly different (FDR <0.05) between current smokers and never smokers. Three metabolites (PC ae C34:3, PC aa C36:1 and glutamate) were identified in both men and women showing the same pattern of variation (higher or lower) (Table 2). Compared with former smokers and never smokers, in male current smokers, the concentrations of four unsaturated diacyl-PCs (PC aa C34:1, PC aa C36:1, PC aa C38:3 and PC aa C40:4) and five amino acids (arginine, aspartate, glutamate, ornithine and serine)

were higher, whereas three saturated diacyl-PCs, one lyso-PC and four acyl-alkyl-PCs, as well as kynurenine showed lower concentrations. In female current smokers, higher levels of carnitine and PC aa C32:1, and lower level of hydroxysphingomyeline (SM (OH)) C22:2 were identified.

Table 2. Smoking-related metabolites in the KORA S4

Results of pair wise comparison by logistic regression of metabolites on smoking status adjusted for age, BMI, and alcohol consumption. Men and women were analysed separately. All results were presented with a false discovery rate (FDR) below 0.05 (in the comparison between current smokers and never smokers, the FDR was calculated by p-value adjusted for all 140 metabolites; for current smokers vs. former smokers and former smokers vs. never smokers, the FDR was calculated by p-value adjusted for the number of metabolites significantly different between current smokers and never smokers). Smoking-related metabolites found in both men and women are in bold. CS: current smokers; FS: former smokers; NS: never smokers; C0: carnitine; PC: phosphatidylcholine; aa: diacyl-; ae: acyl-alkyl-; lyso-PC: acyl-phosphatidylcholine; SM (OH): hydroxysphingomyeline.

Metabolites	**CS vs. NS**		**CS vs. FS**		**FS vs. NS**	
	Odds Ratio (95% CI)	**P-value**	**Odds Ratio (95% CI)**	**P-value**	**Odds Ratio (95% CI)**	**P-value**
Men	**(125 vs. 200)**		**(125 vs. 321)**		**(321 vs. 200)**	
Amino acids						
Arginine	1.7 (1.3, 2.2)	2.6E-05*	1.3 (1.0, 1.6)	0.03*	1.2 (1.0, 1.5)	0.03
Aspartate	1.6 (1.2, 2.0)	2.5E-04*	1.4 (1.1, 1.7)	4.7E-03*	1.1 (0.9, 1.3)	0.36
Glutamate	1.6 (1.2, 2.0)	6.2E-04*	1.4 (1.1, 1.9)	0.02*	1.0 (0.8, 1.3)	0.88
Ornithine	1.4 (1.2, 1.9)	2.2E-03*	1.3 (1.1, 1.7)	8.3E-03*	1.0 (0.9, 1.2)	0.78
Serine	1.4 (1.1, 1.8)	3.5E-03*	1.2 (1.0, 1.5)	0.12	1.1 (0.9, 1.4)	0.25
Kynurenine	0.6 (0.5, 0.9)	3.2E-03*	0.7 (0.5, 0.9)	2.3E-03*	1.0 (0.8, 1.2)	0.88
Phosphatidylcholines						
PC aa C32:3	0.7 (0.5, 0.9)	6.4E-03*	0.8 (0.6, 1.0)	0.07	0.9 (0.7, 1.0)	0.12
PC aa C34:1	1.7 (1.3, 2.2)	2.0E-04*	1.7 (1.3, 2.2)	2.5E-05*	0.9 (0.8, 1.1)	0.49
PC aa C36:0	0.6 (0.5, 0.8)	3.5E-04*	0.6 (0.5, 0.8)	2.7E-04*	1.0 (0.8, 1.2)	0.72
PC aa C36:1	1.6 (1.2, 2.0)	9.4E-04*	1.6 (1.3, 2.0)	8.2E-05*	0.9 (0.8, 1.1)	0.33
PC aa C38:0	0.7 (0.5, 0.9)	2.1E-03*	0.6 (0.5, 0.8)	1.2E-04*	1.0 (0.9, 1.3)	0.64
PC aa C38:3	1.5 (1.1, 1.9)	3.4E-03*	1.3 (1.1, 1.7)	0.01*	1.0 (0.8, 1.2)	0.85
PC aa C40:4	1.5 (1.2, 2.0)	3.4E-03*	1.4 (1.1, 1.8)	3.6E-03*	1.0 (0.8, 1.2)	0.86
PC ae C34:3	0.5 (0.4, 0.7)	3.3E-06*	0.6 (0.5, 0.8)	6.0E-05*	0.9 (0.7, 1.1)	0.23
PC ae C38:0	0.7 (0.5, 0.9)	2.1E-03*	0.6 (0.5, 0.8)	6.7E-04*	1.0 (0.8, 1.2)	0.94
PC ae C38:6	0.7 (0.5, 0.9)	4.8E-03*	0.7 (0.5, 0.8)	6.6E-04*	1.0 (0.8, 1.2)	0.97
PC ae C40:6	0.6 (0.5, 0.8)	8.8E-04*	0.7 (0.5, 0.8)	8.9E-04*	0.9 (0.8, 1.1)	0.33

Lyso-Phosphatidylcholines						
LPC (18:2)	0.7 (0.5, 0.9)	3.3E-03*	0.8 (0.6, 0.9)	0.046*	0.9 (0.7, 1.1)	0.23
Women	**(70 vs. 395)**		**(70 vs. 130)**		**(130 vs. 395)**	
Acylcarnitines						
C0	1.8 (1.4, 2.4)	4.3E-05*	1.5 (1.1, 2.1)	0.01*	1.1 (0.9, 1.4)	0.32
Amino Acid						
Glutamate	1.7 (1.3, 2.2)	1.2E-04*	1.8 (1.3, 2.5)	1.1E-03*	0.9 (0.7, 1.1)	0.17
Phosphatidylcholines						
PC aa C32:1	1.5 (1.1, 1.9)	2.1E-03*	1.4 (1.0, 2.0)	0.03*	1.1 (0.9, 1.4)	0.24
PC aa C36:1	1.6 (1.2, 2.0)	1.1E-03*	1.5 (1.1, 2.0)	0.02*	1.0 (0.8, 1.2)	0.87
PC ae C34:3	0.6 (0.4, 0.8)	7.7E-04*	0.6 (0.4, 0.8)	2.5E-03*	1.0 (0.8, 1.2)	0.94
Sphingomyelines						
SM OH C22:2	0.6 (0.5, 0.8)	2.1E-03*	0.6 (0.4, 0.9)	4.9E-03*	0.9 (0.7, 1.1)	0.35

Among the 21 smoking-related metabolites (18 in men and six in women), 19 were found to be reversible (that is, significant difference between former smokers and current smokers but without significant difference between former smokers and never smokers; FDR <0.05). No irreversible metabolite was observed (that is, significant difference between former smokers and never smokers). Serine and PC aa C32:3 in men were not classified because their concentrations were not significantly different between current smokers and former smokers or between former smokers and never smokers (Table 2). A heat map representing the concentration profiles of the 21 identified metabolites in current smokers, former smokers and never smokers is shown in Figure 4, demonstrating the reversibility of metabolites after smoking cessation.

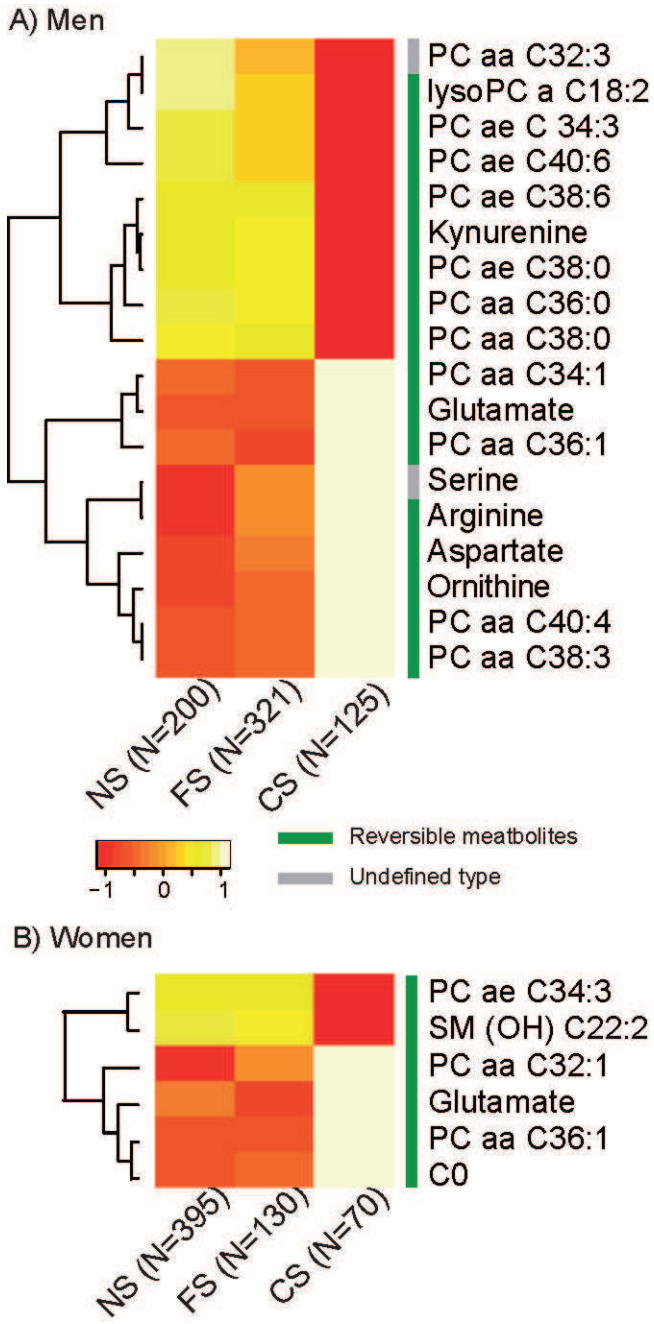

Figure 4. Heat maps of smoking-related metabolites in (A) men and (B) women**.**

The heat map shows mean residues of smoking-related metabolites in current smokers, former smokers and never smokers and the reversibility after smoking cessation. The color of each cell in the heat map represents the relative mean concentration of each metabolite in never smokers, former smokers or current smokers. The number of samples in each group is provided. The bar besides the metabolite names indicates the reversibility of these metabolites after smoking cessation. aa: diacyl-; ae: acyl-alkyl-; C0: carnitine; current smokers: current smokers; FS: former smokers; lyso-PC: acyl-phosphatidylcholine; NS: never smokers; PC: phosphatidylcholine; SM (OH): hydroxysphingomyeline.

In women, SM (OH) C22:2 was significantly associated with cessation time (FDR <0.05); however, there was no such significant metabolite in men (Table 3), indicating a non-linear relationship between cessation time and the reversion of metabolite profile. In addition, the former smokers were grouped by stratified cessation years (0 to 10, 11 to 20, 21 to 30, 31 to 40, over 40 years). For some metabolites (for example, PC ae C38:0, PC aa C36:0 and ornithine), the greatest change of concentration occurred within the first 10 years of cessation compared with current smokers (Figure 5).

Table 3. Cessation time-related metabolites in former smokers

FDR was calculated by *P*-value adjusted for the number of smoking-related metabolites with Benjamini-Hochberg method. aa: diacyl-; ae: acyl-alkyl-; C0: carnitine; FS: former smokers; lyso-PC: acyl-phosphatidylcholine; PC: phosphatidylcholine; SM (OH): hydroxysphingomyeline.

Metabolites	**β Estimate (95% CI)*10^{-3}**	**Pr(>|t|)**	**FDR**
Men			
Arginine	-1.7 (-4.2,0.8)	0.19	0.43
Asparate	-3.4 (-7.5,0.7)	0.1	0.32
Glutamate	-3.9 (-8.4,0.5)	0.09	0.48
Ornithine	-1.3 (-4.1,1.4)	0.34	0.49
Kynurenine	-0.2 (-3.0,2.6)	0.87	0.93
PC aa C34:1	-0.4 (-2.8,2.0)	0.72	0.89
PC aa C36:0	1.9 (-1.2,5.0)	0.24	0.43
PC aa C36:1	-0.4 (-3.3,2.5)	0.77	0.88
PC aa C38:0	2.0 (-1.1,5.1)	0.21	0.42
PC aa C38:3	-1.3 (-4.1,1.4)	0.35	0.47
PC aa C40:4	-2.8 (-6.0,0.5)	0.10	0.32
PC ae C34:3	0.1 (-3.2,3.4)	0.97	0.97
PC ae C38:0	3.0 (-0.4,6.4)	0.08	0.64
PC ae C38:6	1.6 (-1.2,4.4)	0.27	0.43
PC ae C40:6	3.0 (0.3,5.7)	0.03	0.48
LPC(18:2)	2.6 (-0.8,6.1)	0.14	0.37
Women			
C0	1.0 (-3.7,5.8)	0.66	0.66
PC aa C32:1	-6.1 (-13.7,1.6)	0.12	0.31
PC aa C36:1	-2.1 (-7.5,3.3)	0.44	0.52
PC ae C34:3	2.2 (-2.9,7.2)	0.41	0.68
SM OH C22:2*	5.4 (1.8,9.0)	3.70E-03	0.02
Glutamate	3.1 (-4.6,10.8)	0.43	0.64

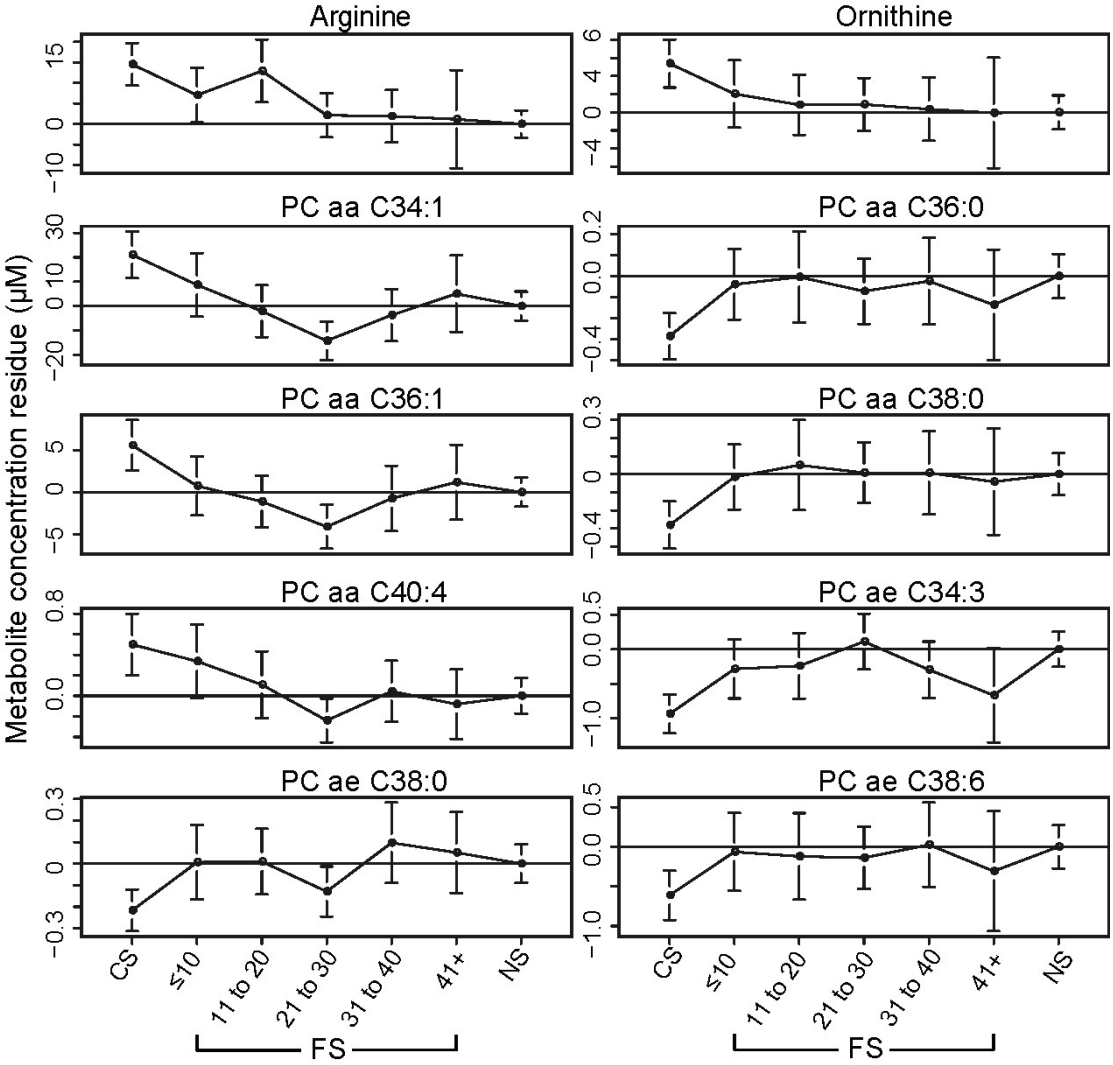

Figure 5. Metabolite concentration variations in relation to smoking cessation time.

Taking never smokers as baseline, figures show the mean residuals of metabolites in different groups of current smokers and former smokers, giving the trend of metabolite variation with cessation time. Former smokers were grouped by stratified cessation time (≤10, 11 to 20, 21 to 30, 31 to 40, 41+). Residuals were calculated by linear regression model (regression of metabolite concentration on age, body mass index and alcohol consumption). aa: diacyl-; ae: acyl-alkyl-; CS: current smokers; FS: former smokers; NS: never smokers; PC: phosphatidylcholine.

Within current smokers, kynurenine and PC ae C34:3, PC ae C38:0 and PC ae C38:6 in men, and PC aa C36:1 in women showed significant association with pack years. In the linear regression model, pack years showed a negative relation (parameter

estimation β <0) to these five metabolites (Table 4) (for example, one pack year increase will lead to a decrease of the kynurenine level in current smokers by 0.33%).

Table 4. Association between smoking intensity (pack years) and metabolites.

Results of linear regression of smoking intensity (pack years) on metabolite concentrations in men and women, adjusted for age, BMI, and alcohol consumption. All smoking-related metabolites presented in Table 3 are listed (*p-value≤0.05). C0: carnitine; PC: phosphatidylcholine; aa: diacyl-; ae: acyl-alkyl-; lyso-PC: acyl-phosphatidylcholine; SM (OH): hydroxysphingomyeline; CI, confidence interval

Metabolites	**β Estimate of pack year (95% CI)*10^{-3}**	**Δ (%)**	**P-value**
Men			
Arginine	-1.1 (-3.6, 1.4)	-0.11%	0.38
Aspartate	2.9 (-1.4, 7.1)	0.29%	0.20
Glutamate	2.9 (-1.2, 6.9)	0.29%	0.17
Ornithine	-2.4 (-5.2, 0.3)	-0.24%	0.09
Serine	1.1 (-1.3, 3.6)	0.11%	0.37
Kynurenine*	-3.3 (-6.1, -0.5)	-0.33%	0.02
PC aa C32:3	-1.4 (-4.3, 1.4)	-0.14%	0.33
PC aa C34:1	-0.9 (-3.5, 1.6)	-0.09%	0.48
PC aa C36:0	-2.3 (-4.9, 0.4)	-0.23%	0.09
PC aa C36:1	-1.4 (-4.6, 1.8)	-0.14%	0.39
PC aa C38:0	-2.1 (-4.9, 0.7)	-0.21%	0.15
PC aa C38:3	1.2 (-1.7, 4.1)	0.12%	0.43
PC aa C40:4	1.3 (-2.5, 5.1)	0.13%	0.51
PC ae C34:3*	-3.7 (-6.4, -0.9)	-0.37%	0.01
PC ae C38:0*	-3.6 (-6.6, -0.5)	-0.36%	0.02
PC ae C38:6*	-2.6 (-5.1, -0.1)	-0.26%	0.04
PC ae C40:6	-1.7 (-4.4, 1.0)	-0.17%	0.22
LPC (18:2)	-3.1 (-6.5, 0.3)	-0.31%	0.07
Women			
C0	1.1 (-4.3, 6.5)	0.11%	0.70
PC aa C32:1	0.2 (-10.5, 10.9)	0.02%	0.97
PC aa C36:1*	6.9 (0.6, 13.2)	0.69%	0.04
PC ae C34:3	-2.7 (-7.7, 2.2)	-0.27%	0.54
SM (OH) C22:2	-2.8 (-7.7, 2.2)	-0.28%	0.28
Glutamate	2.2 (-7.8, 12.2)	0.22%	0.67

3.1.3 Prospective change of metabolite profiles

Among the 89 smokers who had metabolite profiles available in both S4 study and F4, 49 quitted smoking during the 7 years follow up period (quitters). In addition, 432 never smokers in both S4 and F4 had metabolite profile available in both data set (Table 5).

Table 5. Characteristics of the prospective dataset (KORA S4 → F4)

Population characteristics were calculated based on 207 men and 314 women who participated in both KORA S4 and F4 study. Values are provided as mean ± standard deviation (SD). CS: current smokers; FS: former smokers; NS: never smokers; BMI: body mass index. Explanation of variables: Alcohol, alcohol consumption, in g/day; BMI, body mass index, in kg/m^2.

	CS	Quitter	NS
Men(N=207)			
N (%)	31 (15.0%)	30 (14.5%)	146 (70.5%)
Age at S4 (years)	60.2 ± 5.3	63.0 ± 5.0	63.0 ± 5.5
Alcohol(S4/F4)(g/day)	27.7±28.2/20.4±28.7	29.6±31.6/19.3±21.1	22.2±22.8/20.2±19.5
BMI (S4/F4) (kg/m^2)	26.8±2.9 / 26.9±3.3	28.5±3.8 / 28.9±3.9	27.6±3.3 / 27.8±3.4
Women (N=314)			
N (%)	18 (5.7%)	10 (3.2%)	286 (91.1%)
Age at S4	61.0 ± 5.1	59.5 ± 3.1	63.6 ± 5.1
Alcohol(S4/F4)(g/day)	7.6± 11.6/ 7.4 ± 11.8	4.7 ± 6.7/10.7 ± 14.1	7.6± 11.2/7.3 ± 11.4
BMI (S4/F4) (kg/m^2)	27.9± 5.1/ 27.7 ± 5.3	26.9 ± 3.9/27.4 ± 5.1	28.6 ± 4.5/28.9 ± 4.7

Among the 16 reversible metabolites in men, 13 (except kynurenine, glutamate and aspartate) were also measured in KORA F4 using a different kit (see Methods). The linear mixed effect model was applied to investigate the effects of smoking cessation on metabolite concentrations. Among these 13 metabolites, 10 metabolites showed a significant variation in quitters, with a period of smoking cessation from one to seven years, which indicated a reverting process. The arginine level decreased by 11.3% and ornithine by 14.8% in quitters compared with current smokers, whereas PC aa C36:0 increased by 18.5%. Figure 6 shows the prospective changes of the significant metabolites. For women, the same analysis was conducted. Because the number of female quitters was small (N = 10), five metabolites that were measured in both KORA S4 and F4 showed borderline significance (P <0.05). However, none of these metabolites was found to be significant considering FDR <0.05 (see Table 6).

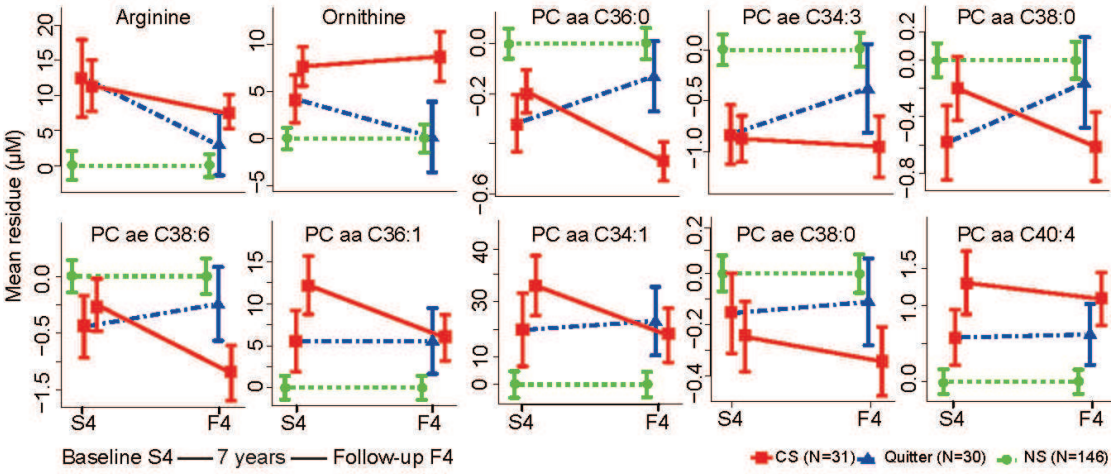

Figure 6. Changes of smoking-related metabolites in current, former and never smokers in the KORA S4 → F4.

Taking the never smokers as baseline, the concentration change of each metabolite is shown as the adjusted mean residue in KORA S4 and F4 in all three groups (CS, FS and NS). Only metabolites with significant prospective change in KORA S4 → F4 are shown in the figure. Residuals were calculated from a linear regression model (regression of metabolite concentration on age, body mass index and alcohol consumption). aa: diacyl-; ae: acyl-alkyl-; CS: current smokers; FS: former smokers; lyso-PC: acyl-phosphatidylcholine; NS: never smokers; PC: phosphatidylcholine.

Table 6. Association of reversible metabolites with smoking status change in the prospective dataset (KORA S4 → F4)

Result of smoking status on metabolite concentrations using linear mixed model for S4 → F4 longitudinal data, adjusted for age, BMI, and alcohol consumption. PC: phosphatidylcholine; aa: diacyl-; ae: acyl-alkyl-; lyso-PC: acyl-phosphatidylcholine; SM (OH): hydroxysphingomyeline.

	β estimate of smoking status (95% CI)	***P***
Men		
Arginine	-0.12 (-0.18, -0.06)	1.4E-04[a]
Ornithine	-0.16 (-0.24, -0.08)	2.1E-04[a]
PC aa C34:1	-0.09 (-0.15, -0.03)	3.3E-03[a]
PC aa C36:0	0.17 (0.09, 0.25)	6.4E-05[a]
PC aa C36:1	-0.12 (-0.18, -0.05)	8.5E-04[a]
PC aa C38:0	0.14 (0.06, 0.22)	3.0E-04[a]
PC aa C38:3	-0.04 (-0.11, 0.02)	1.7E-01
PC aa C40:4	-0.11 (-0.18, -0.03)	6.0E-03
PC ae C34:3	0.14 (0.06, 0.21)	3.5E-04[a]
PC ae C38:0	0.13 (0.05, 0.21)	1.8E-03[a]
PC ae C38:6	0.11 (0.04, 0.18)	1.5E-03[a]

PC ae C40:6	0.08 (0.01, 0.15)	2.1E-02
LPC(18:2)	0.03 (-0.06, 0.11)	5.2E-01
Women		
Carnitine	-0.12 (-0.20, -0.05)	1.4E-03
PC aa C32:1	-0.18 (-0.32, -0.03)	2.1E-03
PC aa C36:1	-0.11 (-0.20, -0.02)	2.0E-02
PC ae C34:3	0.09 (-0.02, 0.19)	0.95
SM (OH) C22.2	0.12 (0.02, 0.22)	1.9E-02

[a] FDR<0.05

3.1.4 Smoking effects on metabolic network

Enrichment analysis of the 21 identified smoking-related metabolites on KEGG pathways showed enrichment in a set of amino acid and lipid metabolism pathways (ether lipid, glycerophospholipid, arginine and proline metabolism). In addition, the impact of the smoking-related metabolites was analyzed in each pathway by measuring their structural importance (see Methods). These metabolites had high betweenness centrality and a strong impact on the enriched pathways (Figure 7 and Table 7).

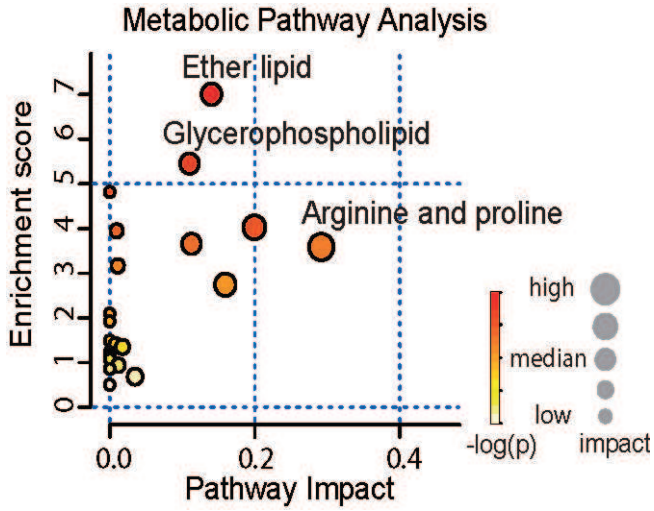

Figure 7. Pathway analyses of smoking-related metabolites

Figure shows enrichment and impact of smoking-related metabolites in Kyoto Encyclopedia of Genes and Genomes pathways. The enrichment scores are shown on y-axis, which was

calculated as the negative logarithm of the *P*-value from an enrichment test. The x-axis indicates the structural impact with a score from 0 to 1 of the smoking-related metabolites in the enriched pathways.

Table 7. Enrichment and impact of smoking-related metabolites in the pathways

Table shows the enrichment and impact scores of smoking-related metabolites in KEGG pathways. The pathway analysis consists of enrichment and a structural impact analysis both based on KEGG database. The -log (p) was considered as the enrichment score. Impact, scored between 0 and 1, indicated the pathway topological importance of the metabolites. In particular, the parameter Total is the total number of compounds in the pathway; the parameter Hits is the actually number of metabolites with significant variations in the pathway; the Raw p is the original p-value calculated from the enrichment analysis; the FDR are calculated by the p values adjusted using Benjamini-Hochberg method.

	Total	**Hits**	**Raw p**	**-log(p)**	**FDR**	**Impact**
Ether lipid metabolism	23	3	9.1E-04	7.01	0.04	0.14
Glycerophospholipid metabolism	39	3	4.3E-03	5.45	0.09	0.11
Cyanoamino acid metabolism	16	2	0.01	4.82	0.13	0
Alanine, aspartate and glutamate metabolism	24	2	0.02	4.03	0.22	0.20
Sphingolipid metabolism	25	2	0.02	3.95	0.22	0.01
Aminoacyl-tRNA biosynthesis	75	3	0.03	3.66	0.25	0.11
Arginine and proline metabolism	77	3	0.03	3.59	0.25	0.29
Glutathione metabolism	38	2	0.04	3.17	0.34	0.01
Glycine, serine and threonine metabolism	48	2	0.06	2.74	0.47	0.16
Linoleic acid metabolism	15	1	0.12	2.09	0.82	0
Sulfur metabolism	18	1	0.15	1.92	0.90	0
alpha-Linolenic acid metabolism	29	1	0.23	1.49	1	0
Vitamin B6 metabolism	32	1	0.25	1.40	1	0.01
Methane metabolism	34	1	0.26	1.35	1	0.02
Nitrogen metabolism	39	1	0.29	1.23	1	0
Butanoate metabolism	40	1	0.30	1.21	1	0
Histidine metabolism	44	1	0.32	1.13	1	0.00
Lysine degradation	47	1	0.34	1.08	1	0
Cysteine and methionine metabolism	56	1	0.39	0.94	1	0.01
Arachidonic acid metabolism	62	1	0.42	0.86	1	0
Tryptophan metabolism	79	1	0.51	0.68	1	0.03
Porphyrin and chlorophyll metabolism	104	1	0.61	0.50	1	0

To systematically investigate how the effects of smoking propagate over the metabolic networks, the association between 175 smoking-related genes were evaluated, previously reported (Beane *et al*, 2007), and the 21 smoking-related metabolites found in this study by analyzing protein-metabolite networks (see Methods). In men, 15 metabolites (LPC(18:2), PC aa C32:3,PC aa C34:1, PC aa C36:0, PC aa C36:1, PC aa C38:0, PC aa C38:3, PC aa C40:4, PC ae C34:3, PC ae

C38:0, PC ae C38:6, PC ae C40:6, arginine, glutamate and serine) were found to be linked with 11 genes (ADH7, AKR1B1, DHRS3, FTL, GALE, GPC1, KRAS, S100A10, SLC7A11, SULF1, PLA2G10) by related enzymes. In women, four metabolites (PC aa C36:1, PC ae C34:3, PC aa C32:1 and glutamate) were closely linked with nine genes (ADH7, AKR1B1, DHRS3, FTL, GALE, GPC1, S100A10, SULF1, PLA2G10) (Figure 8 and Table 8). Similar to enrichment analysis, the network in men and in women could be generally divided into glycerophospholipids and tightly associated proteins as well as amino acids and the associated genes and enzymes. The description of the protein-metabolite and protein-protein interactions was listed in Table 8.

The smoking effects on the networks were reversible. On the aspect of gene expressions, with exception of *SULF1* and *PLA2G10*, all the others in the networks were reversible after smoking cessation (Beane *et al*, 2007). All the metabolites in the network were also reversible, except serine.

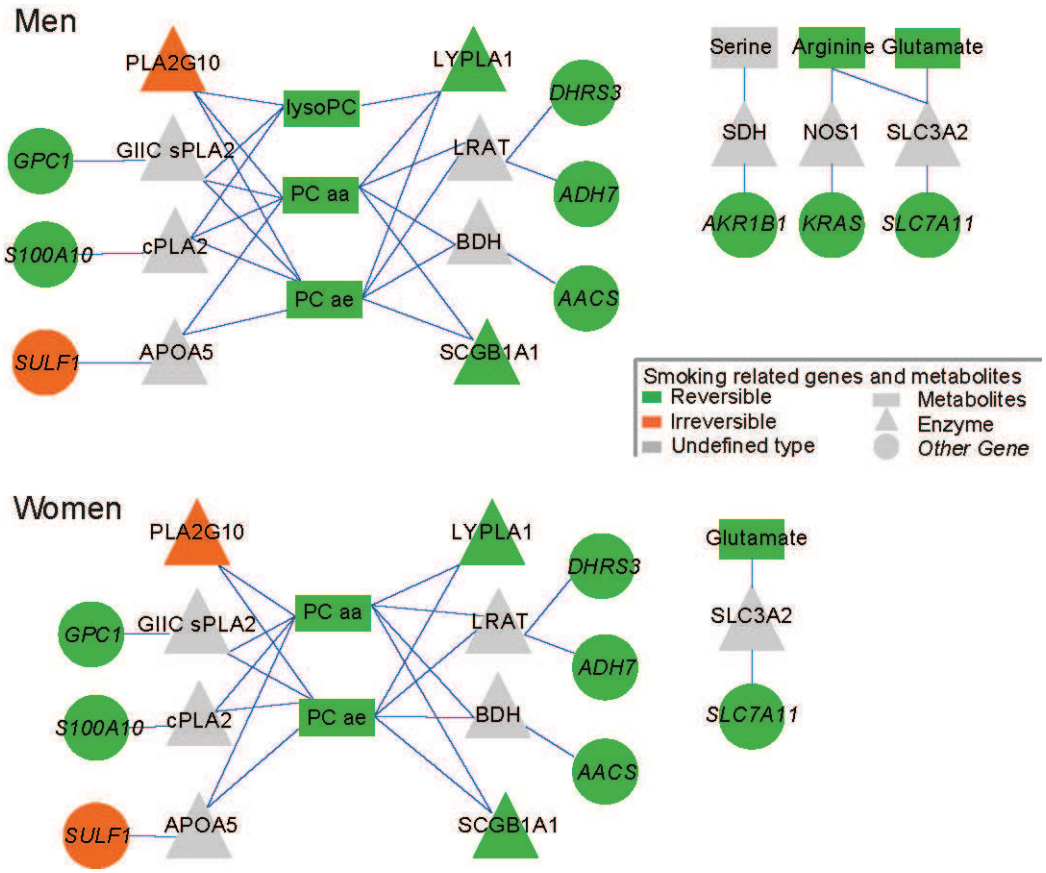

Figure 8. Protein-metabolite networks and pathways of the smoking-related metabolites and genes

Network linking metabolites and proteins encoded by smoking-related genes with maximum one intermediate. Node color indicates the reversibility after smoking cessation. aa: diacyl-; ae: acyl-alkyl-; APOA5: apolipoprotein A-V; BDH: 3-hydroxybutyrate dehydrogenase, type 1; cPLA2: cytosolic phospholipase A2; CS: current smokers; FS: former smokers; GIIC sPLA2: phospholipase A2, membrane associated; LRAT: lecithin retinol acyltransferase; LYPLA1: lysophospholipase I; lyso-PC: acyl-phosphatidylcholine; NOS1: nitric oxide synthase 1; NS: never smokers; PC: phosphatidylcholine; PLA2G10: group 10 secretory phospholipase A2; SCGB1A1: uteroglobin; SDH: serine dehydratase; SLC3A2: solute carrier family 3 member 2.

Table 8. Links between smoking-related metabolites, enzymes and genes

The table describes the links showed in figure 6 of the main text. The smoking related metabolites, enzymes and genes are listed in the first and second columns. The score of interaction is given according to the definition by STRING [1]. A reference for each link and a short description is provided. The Column of reaction shows the possible biochemical reaction of the corresponding link or the type of protein interaction. The enzymes includes, phospholipase A2, membrane associated (GIIC sPLA2), cytosolic phospholipase A2 (cPLA2), group 10 secretory phospholipase A2 (PLA2G10), lysophospholipase I (LYPLA1), apolipoprotein A-V (APOA5), uteroglobin (SCGB1A1), lecithin retinol acyltransferase (LRAT), nitric oxide synthase 1 (NOS1), solute carrier family 3 member 2 (SLC3A2), serine dehydratase (SDH), 3-hydroxybutyrate dehydrogenase, type 1 (BDH). The smoking related gene/protein includes, S100 calcium binding protein A10 (*S100A10*), glypican 1 (*GPC1*), sulfatase 1 (*SULF1*), alcohol dehydrogenase 7 (*ADH7*), dehydrogenase member 3 (*DHRS3*), aldose reductase (*AKR1B1*), acetoacetyl-CoA synthetase (*AACS*), V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*), solute carrier family 7 (*SLC7A11*) and three enzyme listed above, PLA2G10, LYPLA1, SCGB1A1. The links in the network for male and female CS are combined and listed together. Smoking-related genes are show in italic. C0: carnitine; PC: phosphatidylcholine; aa: diacyl-; ae: acyl-alkyl-; lyso-PC: acyl-phosphatidylcholine; SM (OH): hydroxysphingomyeline

Proteins/Enzymes with alternative abbreviations:

PLA2G2A (GIIC sPLA2), Phospholipase A2, membrane associated"; PLA2G4A (cPLA2), Cytosolic phospholipase A2; BDH1 (BDH), D-beta-hydroxybutyrate dehydrogenase, mitochondrial; SDS (SDH), L-serine dehydratase

Metabolites / Enzymes	**Proteins/ Smoking related gene**	**Score**	**Type**	**Description**	**Reaction**
Glutamate	SLC3A2	1	catalytic	SLC7A11-mediated exchange of extracellular cysteine and cytosolic glutamate [1]	transport
Arginine	SLC3A2	1	catalytic	SLC7A7 (y+LAT1)-mediated exchange of extracellular leucine for cytosolic arginine [2]	transport
Arginine	NOS1	1	oxidoreductase	Arginine and proline metabolism [2]	L-arginine + n NADPH + n H+ + m O2 = citrulline + nitric oxide + n NADP+

PC aa/ae CX:Y, LPC(18:2)	*PLA2G10*	1	phospholipase	Glycerophospholipid metabolism [3]	phosphatidylcholine + H2O = 1-acylglycerophosphocholine + a carboxylate
PC aa/ae CX:Y, LPC(18:2)	cPLA2	1	phospholipase	Glycerophospholipid metabolism [3]	phosphatidylcholine + H2O = 1-acylglycerophosphocholine + a carboxylate
PC aa/ae CX:Y, LPC(18:2)	*PLA2G2A*	1	phospholipase	Glycerophospholipid metabolism [3]	phosphatidylcholine + H2O = 1-acylglycerophosphocholine + a carboxylate
PC aa/ae CX:Y, LPC(18:2)	*LYPLA1*	1	hydrolase	Hydrolyzes fatty acids from S-acylated cysteine residues in proteins such as trimeric G alpha proteins or HRAS. Has depalmitoylating activity and also low lysophospholipase activity [4]	2-lysophosphatidylcholine + H2O=glycerophosphocholine + a carboxylate
PC aa/ae CX:Y	APOA5	1	lipid binding	Interact with phosphatidylcholine via lipoprotein lipase (LPL)[5, 6]	
PC aa/ae CX:Y	LRAT	1	phosphatidylcholine-retinol O-acyltransfera	Transfers the acyl group from the sn-1 position of phosphatidylcholine to all-trans retinol, producing all-trans retinyl esters. Retinyl esters are storage forms of vitamin A. [7]	phosphatidylcholine + retinol---[cellular-retinol-binding-protein] = 2-acylglycerophosphocholine + retinyl-ester---[cellular-retinol-binding-protein]
PC aa/ae CX:Y	*SCGB1A1*	1	binding	Binds phosphatidylcholine, potent inhibitor of phospholipase A2 [8-10]	
PC aa/ae CX:Y	BDH	1	activation	BDH activated by phosphotadylcholine	
Serine	SDS	1	catalytic	Binding and dehydrate [11]	L-threonine = 2-oxobutanoate + NH3
SM OH C22:2	*SGMS1*	1	catalytic	Bidirectional lipid cholinephosphotransferase capable of converting phosphatidylcholine (PC) and ceramide to sphingomyelin (SM) and diacylglycerol (DAG) and vice versa. [12]	a ceramide + a phosphatidylcholine = a sphingomyelin + a 1,2-diacyl-sn-glycerol

SLC7A11	SLC3A2	0.99	Binding	1. SLC7A11:SLC3A2 heterodimer; SLC7A11-mediated exchange of extracellular cysteine and cytosolic glutamate. [13] 2. In vivo Experimental data	
KRAS	NOS1	0.83	same pathway	Long-term depression [14]	
GIIC sPLA2	*GPC1*	0.92	binding	Inferred from physical interaction [15]	
cPLA2	*S100A10*	0.83	Inhibition	The antiinflammatory protein annexin-1 (ANXA1) and the adaptor S100A10 (p11), inhibit cytosolic phospholipase A2 (cPLA2alpha) by direct interaction. [16]	S100A10 --| cPLA2
APOA5	*SULF1*	0.72	Binding	low-density lipoprotein receptor family and glycosylphosphatidylinositol high-density lipoprotein binding protein1. [17]	
BDH	*AACS*	0.96	same pathway	Butanoate metabolism [4]	up and down stream
LRAT	*ADH7*	0.9	same pathway	Retinol metabolism [4]	up and down stream
LRAT	*DHRS3*	0.96	same pathway	Retinol metabolism [4, 18]	up and down stream
SDH	*AKR1B1*	0.8	inter pathway	Inter-pathway connection between 'Glycine, serine and threonine metabolism' and 'Pyruvate metabolism' [4]	up and down stream

[1]. Wu, G. and C.J. Meininger, Arginine nutrition and cardiovascular function. J Nutr, 2000. 130(11): p. 2626-9.
[2]. Dijkstra, E.W., A note on two problems in connexion with graphs. Numerische Mathematik, 1959. 1: p. 2.
[3]. Chen, H., et al., Detrimental metabolic effects of combining long-term cigarette smoke exposure and high-fat diet in mice. Am J Physiol Endocrinol Metab, 2007. 293(6): p. E1564-71.

[4]. Kanehisa, M., et al., The KEGG databases at GenomeNet. Nucleic Acids Res, 2002. 30(1): p. 42-6.
[5]. Wishart, D.S., et al., HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res, 2009. 37(Database issue): p. D603-10.
[6]. Wichmann, H.E., C. Gieger, and T. Illig, KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen, 2005. 67 Suppl 1: p. S26-30.
[7]. Jahng, W.J., L.L. Xue, and R.R. Rando, Lecithin retinol acyltransferase is a founder member of a novel family of enzymes. Biochemistry, 2003. 42(44): p. 12805-12812.
[8]. Shen, X.Z., et al., Hormonal regulation of rabbit uteroglobin gene transcription. Endocrinology, 1983. 112(3): p. 871-6.
[9]. Tikkanen, M.J., et al., Effects of oestradiol and levonorgestrel on lipoprotein lipids and postheparin plasma lipase activities in normolipoproteinaemic women. Acta Endocrinol (Copenh), 1982. 99(4): p. 630-5.
[10]. Demel, R.A., K. Shirai, and R.L. Jackson, Lipoprotein lipase-catalyzed hydrolysis of tri[14C]oleoylglycerol in a phospholipid interface. A monolayer study. Biochim Biophys Acta, 1982. 713(3): p. 629-37.
[11]. Tandon, R.S. and K.C. Misra, Threonine and serine dehydratase activity in the buffalo liver-fluke Fasciola indica. J Helminthol, 1980. 54(4): p. 259-62.
[12]. Tani, M. and O. Kuge, Sphingomyelin synthase 2 is palmitoylated at the COOH-terminal tail, which is involved in its localization in plasma membranes. Biochemical and Biophysical Research Communications, 2009. 381(3): p. 328-332.
[13]. Kim, J.Y., et al., Human cystine/glutamate transporter: cDNA cloning and upregulation by oxidative stress in glioma cells. Biochimica Et Biophysica Acta-Biomembranes, 2001. 1512(2): p. 335-344.
[14]. Yun, H.Y., et al., Nitric oxide mediates N-methyl-D-aspartate receptor-induced activation of p21ras. Proc Natl Acad Sci U S A, 1998. 95(10): p. 5773-8.
[15]. Boilard, E., et al., Interaction of low molecular weight group IIA phospholipase A2 with apoptotic human T cells: role of heparan sulfate proteoglycans. FASEB J, 2003. 17(9): p. 1068-80.
[16]. Borot, F., et al., Eicosanoid Release Is Increased by Membrane Destabilization and CFTR Inhibition in Calu-3 Cells. PLoS One, 2009. 4(10).
[17]. Forte, T.M., X. Shu, and R.O. Ryan, The ins (cell) and outs (plasma) of apolipoprotein A-V. Journal of Lipid Research, 2009. 50: p. S150-S155.
[18]. Ashla, A.A., et al., Genetic analysis of expression profile involved in retinoid metabolism in non-alcoholic fatty liver disease. Hepatol Res, 2010. 40(6): p. 594-604

3.2 Effects of smoking on multi-level 'omics' profiles

As shown in the first study, the effects of smoking on gene expression and metabolites were concordant that were both reversible after smoking cessation. In this study, the effects of smoking on multi-level omics profiles were investigated; comparisons were drawn between current smokers and never smokers to explore the variations in epigenomic and transcriptomic profiles, and their relation to the metabolite concentrations. The smoking-associated CpG sites, genes and metabolites were used in the following mediation analysis.

3.2.1 Smoking-associated CpG sites

From the KORA F4, blood samples from 753 never smokers and 262 current smokers were selected for DNA methylation profiling (Table 9). The data which arose from this formed the discovery data set. Additionally, 250 never smokers and current smokers from the KORA F3 were selected to form the replication dataset. In both datasets it was revealed that the current smokers showed a significant higher percentage of alcohol drinkers as well as lower BMI levels. In the KORA F4 data set, the current smokers are younger in age and are more likely to be males.

Table 9. Population characteristics of samples used in the analysis of smoking effect on methylation

	KORA F4			**KORA F3**		
	Never smoker	**Current smoker**	***P***	**Never smoker**	**Current smoker**	**P**
N	753	262		250	250	
Age (years)	62.12(8.98)	56.96(7.01)	9.17E-17	53.15(9.67)	52.62(9.62)	0.49
Sex (female)	493	120	3.21E-08	120	120	1
BMI (kg/m^2)	28.04(4.56)	27.11(4.84)	8.0E-4	27.53(4.49)	26.78(4.57)	0.01
Alcohol (g/day)	11.73(16.79)	18.2(24.33)	0.007	13.82(17.41)	18.41(21.34)	0.08
Alcohol drinker [a] (%)	110 (14.6%)	61(23.3%)	0.002	40 (16%)	72 (28.8%)	8.4E-4
Diabetes (%)	61(8.1%)	19 (7.3%)	0.79	12 (4.8%)	19(7.6%)	0.26

[a] men: > 40g/day; women: > 20g/day;

In total, 641 CpG sites were found to be (Bonferroni threshold, 10^{-7}) associated with smoking in the discovery dataset (Table S2 in the Appendix) upon using the linear regression model which was adjusted for covariates including age, sex, BMI, alcohol consumption (men: > 40g/day; women: > 20g/day) and white blood cell portions. Amongst these CpG sites, 361 sites were also found (Bonferroni threshold, p-value < 10^{-5}) in the replication dataset upon adjustment of the same model (Table S2 in the Appendix).

3.2.2 Smoking-associated differential gene expression

Gene expression profiles from blood samples were available for 490 never smokers and 66 current smokers in KORA F4. In the replication data set F3, expression profiles for 183 never smokers and 45 current smokers were measured (Table 10).

Table 10. Population characteristics of the samples used in the analysis of smoking effect on gene expression

	KORA F4			KORA F3		
	Never smoker	**Current smoker**	***P***	**Never smoker**	**Current smoker**	***P***
N	490	66		183	45	
Age (years)	70.71(5.37)	67.74(5.29)	4.07E-05	66.78(7.05)	62.93(6.83)	0.002
Sex (female)	343	26	2.60E-06	111	22	0.18
BMI (kg/m^2)	28.88(4.56)	27.51(4.87)	0.006	28.43(3.8)	27.89(3.82)	1
Alcohol (g/day)	10.82(15.43)	14.01(19.72)	0.29	12.02(14.1)	14.96(21.62)	0.58
Alcohol drinker[a](%)	71(14.5%)	11(16.7%)	0.58	26 (14.2%)	11 (24.4%)	0.11
Diabetes (%)	53 (10.8%)	11 (16.7%)	0.21	24 (13.1%)	9 (20%)	0.24

[a] men: > 40g/day; women: > 20g/day

In the KORA F4 study, using the linear regression model which was adjusted for age, sex, BMI and alcohol consumption category, 23 gene expressions were found to differ significantly between current smokers and never smokers (corrected for multiple testing, p-value <10^{-6}) (Table 11). Amongst the 23 genes, the expressions of two genes (*LRRN3*, *CLDND1*) were also significantly different between the two groups in

the replication dataset, after correction for multiple testing using the Bonferroni criteria (p-value $<10^{-3}$, Table 11). The inflation coefficients (lambda) were 1.38 and 1.26 respectively in the KORA F4 and the KORA F3 respectively, which indicate no strong potential confounding effects (Figure 9).

Table 11. Differential gene expression between current smokers and never smokers in the KORA F4 and replication in the KORA F3

The expression profile in F4 and F3 were measured using different platforms, including different set of probes for each gene. Blank entries in the table indicate probes not available in the corresponding dataset. The association β (95% confidence interval) were estimated as the regression coefficients in the linear model of metabolite ~ smoking + age + sex + BMI + alcohol consumption, indicating the differences between current smokers and never smokers in SD of log-transformed metabolite concentration. The probes and genes in bold font are significantly associated with smoking in both F4 and F3 studies. CI: confidence interval.

		KORA F4		**KORA F3**	
	Gene	**Association β (95% CI)**	**P**	**Association β (95% CI)**	**P**
ILMN_1718565	*CDKN1C*	-0.33(-0.44,-0.22)	1.32E-08	-0.12(-0.25,0.00)	0.05
ILMN_1710326	***CLDND1***	0.27(0.19,0.36)	3.78E-10	0.27(0.16,0.38)	2.41E-06
ILMN_2352563	*CLDND1*	0.29(0.19,0.39)	4.03E-09		
ILMN_1695991	*COLQ*			-0.17(-0.30,-0.04)	0.01
ILMN_1670872	*COLQ*			0.005(-0.02,0.03)	0.72
ILMN_1734325	*COLQ*			-0.001(-0.03,0.03)	0.93
ILMN_2329114	*COLQ*	-0.26(-0.36,-0.16)	5.19E-07		
ILMN_1656501	*DUSP5*	-0.29(-0.38,-0.20)	1.61E-10	-0.01(-0.07,0.05)	0.71
ILMN_1661248	*EDG8*			-0.01(-0.06,0.04)	0.76
ILMN_2073184	*EDG8*	-0.41(-0.54,-0.27)	4.51E-09		
ILMN_1717902	*ERBB2*			-0.03(-0.06,0.00)	0.02
ILMN_1728761	*ERBB2*			0.01(-0.02,0.04)	0.38
ILMN_1694303	*ERBB2*			-0.02(-0.06,0.02)	0.35
ILMN_2352131	*ERBB2*	-0.26(-0.36,-0.16)	5.57E-07		
ILMN_1761945	*FGFBP2*	-0.42(-0.59,-0.26)	9.30E-07	-0.08(-0.29,0.12)	0.43
ILMN_1692464	*FLJ20699*	-0.36(-0.48,-0.25)	1.89E-09	-0.12(-0.25,0.02)	0.08
ILMN_2059886	*FLJ20699*	-0.30(-0.40,-0.20)	8.36E-09		
ILMN_1738746	*GPR15*			0.02(0.00,0.05)	0.07
ILMN_2192779	*GPR15*	0.22(0.19,0.26)	2.87E-28		
ILMN_1684349	*IL2RB*	-0.28(-0.39,-0.17)	4.59E-07	-0.17(-0.34,0.00)	0.05
ILMN_1773650	***LRRN3***	1.19(1.03,1.35)	5.75E-42	0.51(0.39,0.63)	1.28E-15
ILMN_2048591	*LRRN3*	0.56(0.47,0.65)	7.31E-31		
ILMN_1682993	*NKG7*	-0.38(-0.52,-0.25)	4.62E-08	-0.20(-0.39,0.00)	0.05
ILMN_1802151	*OSBPL5*	-0.33(-0.43,-0.23)	5.03E-10	-0.09(-0.21,0.03)	0.15
ILMN_1671891	*PID1*	0.25(0.16,0.34)	2.29E-07		
ILMN_1813208	*PID1*			0.09(0.02,0.15)	0.01
ILMN_1770800	*PODN*	-0.15(-0.21,-0.10)	2.10E-08	-0.05(-0.08,-0.01)	0.01

ILMN_1740633	*PRF1*	-0.44(-0.58,-0.30)	1.87E-09	-0.19(-0.38,0.00)	0.05
ILMN_1693129	*RNF165*			-0.02(-0.06,0.02)	0.43
ILMN_1792389	*RNF165*	-0.37(-0.50,-0.24)	3.26E-08		
ILMN_1712673	*SASH1*	0.22(0.15,0.29)	2.01E-09	0.02(-0.01,0.05)	0.18
ILMN_2185984	*SASH1*	0.23(0.17,0.29)	9.56E-14		
ILMN_1701237	*SH2D1B*	-0.34(-0.47,-0.21)	3.74E-07	-0.002(-0.11,0.11)	0.97

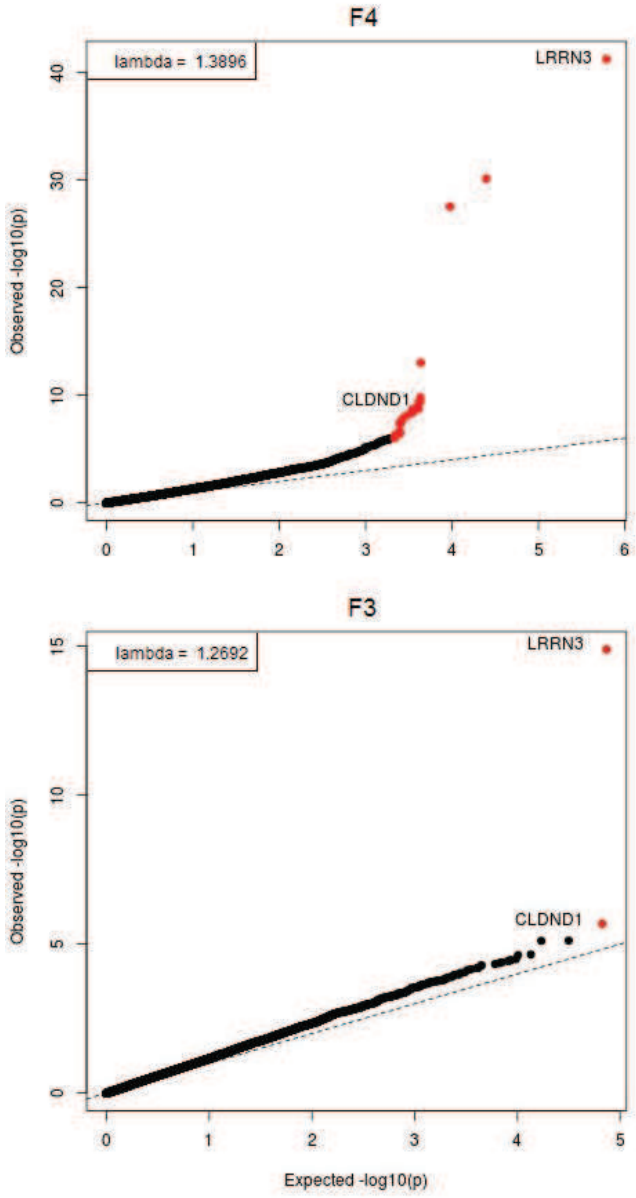

Figure 9. QQ plot for the p-values of the associations between smoking and gene expression Red dot indicates the genes found significantly associated with smoking in F3 and F4.

3.2.3 Smoking-associated metabolites

The 18 metabolites associated with smoking, as reported in section 3.1, were used as the candidate metabolites for the following analysis. These metabolites were: arginine, C0, LPC (18:2), Ornithine, PC aa C32:1, PC aa C32:3, PC aa C34:1, PC aa C36:0, PC aa C36:1, PC aa C38:0, PC aa C38:3, PC aa C40:4, PC ae C34:3, PC ae C38:0, PC ae C38:6, PC ae C40:6, Serine, SM (OH) C22:2 (Xu *et al*, 2013).

3.2.4 Mediation analysis

Mediation analyses were performed using the 18 metabolites, 2 genes, and 361 CpG sites to ultimately unveil their interactions amongst one another in response to smoking. In the KORA F4, 729 samples had all three levels of the 'omics' data, while in the KORA F3 only 57 samples were available for the study (Figure 10). The mediation analysis was used to find the CpG sites and genes which mediate the effects of smoking on metabolite concentrations (smoking → methylation → expression→ metabolites). The mediation effects were analyzed in two steps: 1) mediation of methylation for the association between smoking and gene expression; 2) mediation of expression for the association between methylation and concentrations of metabolites.

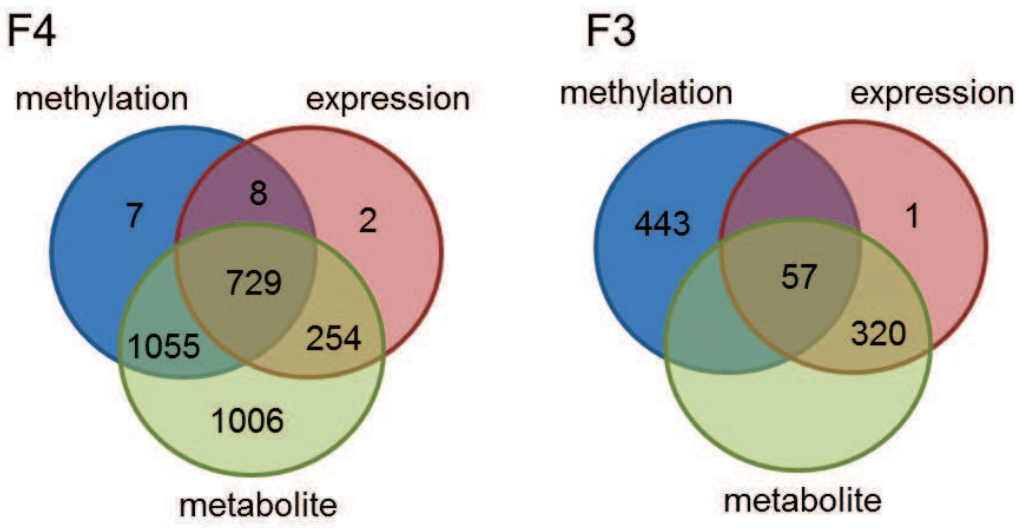

Figure 10. Venn diagram showing the availability of different 'omics' data in the KORA F3 and F3 studies

The available 'omics' data in the KORA F4 and F3 studies are shown in the form of venn diagrams. The number in the overlapping region of the circles indicates the number of samples, which have all types of 'omics' data in the corresponding study.

3.2.4.1 Methylation mediated association between smoking and gene expression (smoking → methylation → expression)

The mediation effect of methylation for the association between smoking and gene expression were tested. Seven CpG sites showed significant mediation effects for the association between smoking and *LRRN3* gene expression in the KORA F4 (significant after Bonferroni correction, p-value <7E-5) (Table 12). However, in the replication dataset (KORA F3), none of the mediation effects were significant (p-value <0.05) (Table 12).

Amongst these CpG sites, cg09837977 has cis-regulation in relation with LRRN3, which sits within the 1Mbp region of the transcription starting site of LRRN3 gene (Figure 11). LRRN3 was significantly up-regulated in current smokers, and cg09837977 showed hypo-methylation in current smokers. Meanwhile, the Pearson correlation also showed a major negative association between LRRN3 gene expression and cg09837977 methylation amongst the whole population (Pearson

correlation coefficients (95%CI) = -0.45 (-0.52,-0.37), p-value = 4.08E-22, Figure 12). The correlation (95% CI) in subgroups are -0.46 (-0.51,-0.39) in never smokers, and -0.42 (-0.62,-0.18) in current smokers.

Table 12. Mediation effect of seven CpG sites for the association between *LRRN3* gene expression and smoking in the KORA F4 and KORA F3

Mediation analysis is conducted using linear regression model adjusted for age, sex, BMI, alcohol consumption, metabolite ~ α'*CpG + γ*expression. α' is indicated as smoking-expression association; γ indicates gene methylation-expression association in the table.

Cpg site	Gene	**Mediation Effect**	***P***	**Smoking-expression α' association (95% CI)**	***P***	**Methylation-expression γ Association (95% CI)**	***P***
F4							
cg23771366	PRSS23	0.11(0.06,0.17)	4.04E-05	1.09(0.9,1.27)	1.80E-28	-0.42(-0.59,-0.25)	1.96E-06
cg00073090		-0.13(-0.19,-0.07)	6.20E-06	1.33(1.15,1.51)	1.27E-41	1.12(0.78,1.46)	1.86E-10
cg02532700	NCF4	-0.12(-0.18,-0.06)	1.67E-05	1.32(1.14,1.5)	4.91E-42	0.38(0.27,0.48)	3.20E-12
cg00501876	CSRNP1	0.15(0.08,0.21)	7.33E-06	1.06(0.88,1.23)	1.49E-29	-0.56(-0.68,-0.44)	3.29E-18
cg14753356		-0.13(-0.19,-0.07)	7.78E-06	1.33(1.15,1.51)	1.25E-40	0.29(0.19,0.39)	3.06E-08
cg26729380	TNF	-0.13(-0.19,-0.07)	4.04E-05	1.33(1.16,1.5)	2.72E-44	0.49(0.39,0.6)	1.56E-18
cg09837977	LRRN3	0.18(0.1,0.26)	6.20E-06	1.02(0.85,1.19)	3.38E-30	-0.45(-0.52,-0.38)	1.37E-32
F3							
cg23771366	PRSS23	-0.06(-0.17,0.05)	0.13	0.73(0.49,0.97)	3.05E-07	0.29(-0.19,0.78)	0.24
cg00073090		-0.04(-0.14,0.07)	0.25	0.7(0.46,0.94)	7.31E-07	0.31(-0.56,1.18)	0.48
cg02532700	NCF4	-0.05(-0.14,0.05)	0.16	0.71(0.48,0.95)	2.71E-07	0.19(-0.16,0.53)	0.29
cg00501876	CSRNP1	-0.01(-0.06,0.05)	0.41	0.67(0.44,0.89)	4.49E-07	0.03(-0.23,0.29)	0.82
cg14753356		-0.05(-0.18,0.08)	0.23	0.71(0.46,0.97)	1.38E-06	0.12(-0.19,0.42)	0.46
cg26729380	TNF	-0.04(-0.11,0.03)	0.15	0.7(0.48,0.93)	1.59E-07	0.25(-0.16,0.65)	0.24
cg09837977	LRRN3	0.02(-0.03,0.06)	0.26	0.65(0.42,0.87)	6.65E-07	-0.07(-0.28,0.13)	0.48

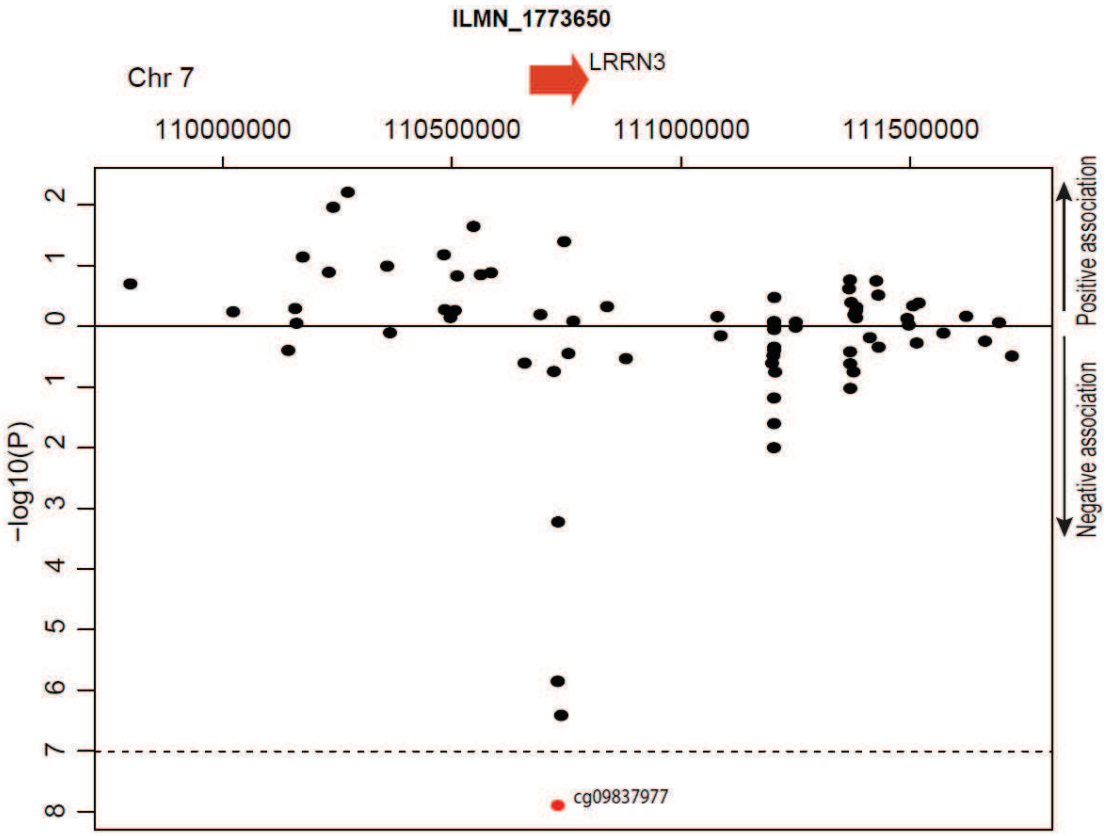

Figure 11. Association between the *LRRN3* gene and the CpG site cg09837977 in the KORA F4

LRRN3 gene expression was indicated by the expression level measured by the probe ILMN_1773650 from the Illumina HT12 v3 expression array. The x axis indicates the position in the chromosome 7. The y axis indicates the –log10 transformed P-value for the association between the CpG sites and *LRRN3* gene expression. CpG sites above 0 are positively associated with LRRN3 gene expression, while negatively associated with smoking if blow 0. The LRRN3 gene is shown as a segment with arrow indicating the direction of transcription. Red color indicates up-regulation of *LRRN3* expression in current smokers. CpG sites measured with in 1Mb region of TSS of *LRRN3* are shown in the figure as points. Red dots indicate CpG sites significantly associated with LRRN3 expression in a genome wide level.

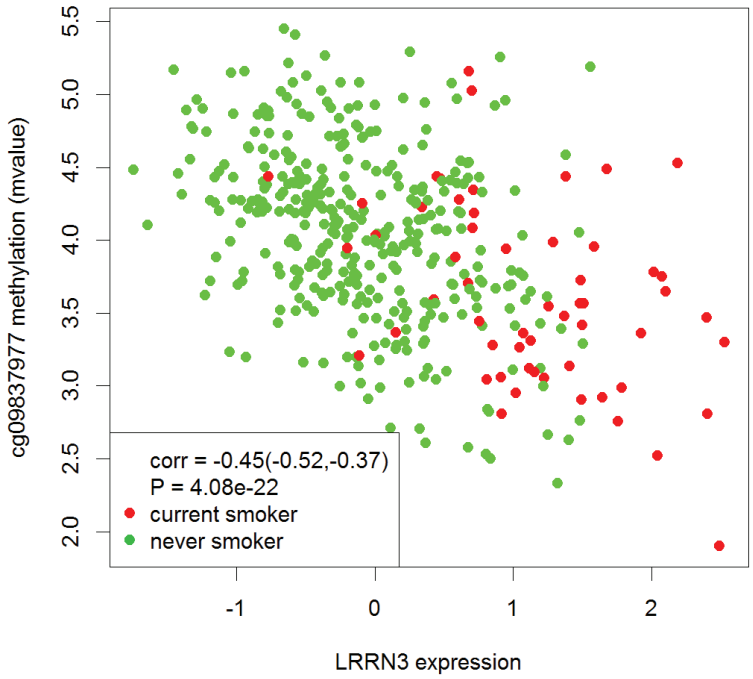

Figure 12. Correlation between *LRRN3* expression and cg09837977 methylation in the KORA F4

Current smokers and never smokers were indicated by points with different colors as shown in the figure legend. Pearson correlation with 95% confidence interval was calculated, and tested for significance using t-test.

The associations between cessation years and gene expression and methylation were assessed in former smokers. The expression and methylation level showed a decreasing trend with increased cessation years (Figure 13). The influence of smoking pack-years in current smokers was analyzed to establish the influence of smoking intensity on the *LRRN3* expression and cg09837997 methylation. However, no significant association was found.

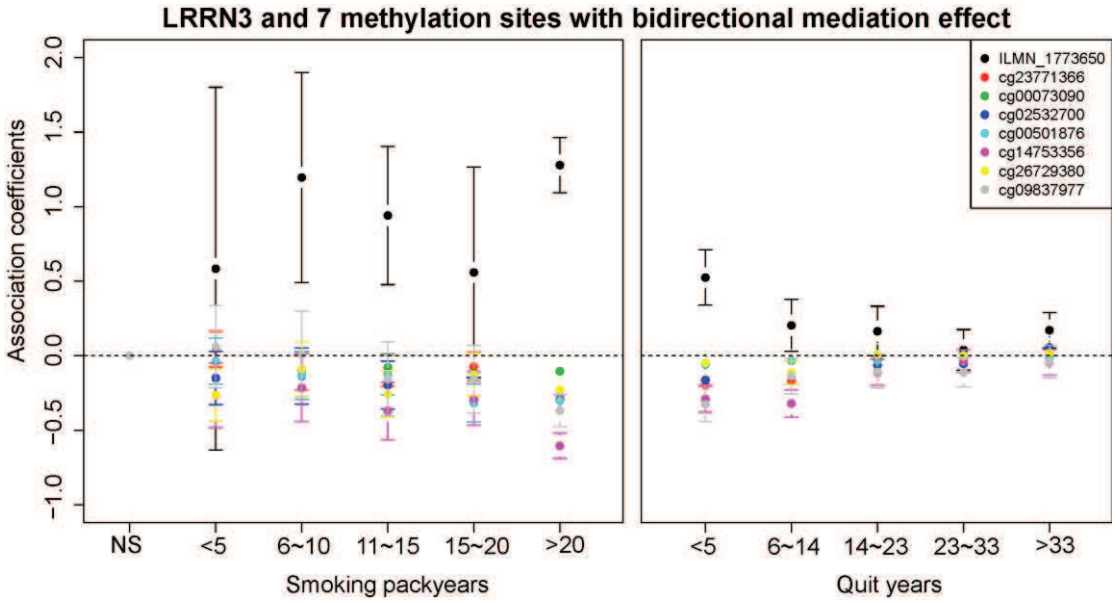

Figure 13. The expression of *LRRN3* and methylations of seven CpG sites at different pack years and cessation years in the KORA F4

3.2.4.1 Gene expression mediated association between methylation and metabolite concentration (methylation → expression → metabolite)

The mediation effects of expressions for the association between methylations and metabolite concentrations were tested. In the KORA F4, 55 smokers and 357 never-smokers were available for the analysis (Table 13).

After adjusting for covariates, 18 genes showed borderline significance (p-value <0.05, without adjustment of multiple testing) for the mediation of the association between CpG sites and metabolites (Table 14). Two CpG sites, cg26729380 and cg14753356, which significantly mediated the association between smoking and *LRRN3* gene expression, were significantly associated with LPC (18:2) and PC ae C34:3 levels. On the other hand, the *LRRN3* expression showed significant mediation effects for the association between the two CpG sites and two metabolites (Table 14). Cg14753356 is positively associated with the LPC (18:2) levels (β (95% CI) = 0.14(-0.01, 0.30), p-value = 0.07), while *LRRN3* expression mediates for about 21% (0.03/0.14*100% ≈ 21%) of the effects for the association. On the contrary, *LRRN3*

attenuated 20% of the association between cg26729380 and PC ae C34:3, which showed a negative mediation effect (Table 14). Cg09837977 was not significantly associated with any smoking related metabolites tested in this study. In the replication dataset, only 28 current smokers and 29 non-smokers had three levels of 'omics' data available for the mediation analysis (Table 13). The mediation effects were not replicated in the KORA F3 study (Table 14).

Table 13. Population characteristics of the datasets for the mediation analysis

	KORA F4			**KORA F3**		
	Never smokers	**Current smokers**	***P***	**Never smokers**	**Current smokers**	***P***
N	357	55		29	28	
Age (years)	69.17(4.37)	66.36(3.89)	1.74E-05	61.45(4.48)	61.57(5.59)	0.92
sex (female)	248	22	5.12E-05	12	15	0.43
BMI (kg/m^2)	28.75(4.41)	27.52(5)	0.02	27.77(3.71)	27.47(3.89)	0.81
alcohol (g/day)	11.56(16.33)	15.04(20.4)	0.37	14.65(17.82)	11.24(16.98)	0.29
alcohol drinker (%)	56 (15.7%)	9 (16.4%)	0.84	5 (17.2%)	6 (21.4%)	0.75
Diabetes (%)	37 (10.4%)	9 (16.4%)	0.25	1 (3.6%)	8 (28.6%)	0.01

Table 14. Mediation effects of gene expression for the association between methylation and metabolite in the KORA F4 and KORA F3

Mediation analysis is conducted using linear regression model adjusted for age, sex, BMI, alcohol consumption, metabolite ~ α'*CpG + γ*expression. α' is indicated as methylation-metabolite association; γ indicates gene expression-metabolite association in the table. The two CpG sites mediated the association between smoking and *LRRN3* expression were shown in bold.

							Methylation-metabolite		Gene expression-metabolite	
CpG sites		**Gene expression**		**Metabolite**	**Mediation Effect**	**P**	**α' (95% CI)**	**P**	**γ (95% CI)**	**P**
F4										
cg00295485	UXS1	ILMN_1773650	LRRN3	LPC (18:2)	-0.03(-0.06,0.00)	0.03	-0.11(-0.23,0.01)	0.08	0.11(0.00,0.22)	0.05
cg00893603	ATP8A2	ILMN_1773650	LRRN3	LPC (18:2)	-0.02(-0.03,0.00)	0.03	-0.09(-0.17,-0.01)	0.02	0.11(0.00,0.22)	0.05
cg01208318		ILMN_1773650	LRRN3	LPC (18:2)	-0.02(-0.04,-0.01)	0.01	0.12(0.05,0.20)	1.80E-03	0.17(0.06,0.27)	0.001
cg01731783	C14orf43	ILMN_1710326	CLDND1	PC ae C38:0	0.04(0.00,0.07)	0.02	-0.35(-0.56,-0.15)	8.40E-04	-0.25(-0.47,-0.03)	0.02
cg04039799	NAV2	ILMN_1773650	LRRN3	PC ae C38:0	-0.04(-0.08,0.00)	0.03	-0.13(-0.30,0.04)	0.14	0.12(0.00,0.24)	0.05
cg04425624	TNF	ILMN_1773650	LRRN3	PC ae C38:0	0.06(0.00,0.11)	0.03	0.21(-0.03,0.44)	0.08	0.12(0.00,0.23)	0.05
cg05824218	RARA	ILMN_1710326	CLDND1	PC ae C38:0	-0.04(-0.07,0.00)	0.02	0.37(0.16,0.57)	4.00E-04	-0.26(-0.48,-0.04)	0.02
cg06235438	ITGAL	ILMN_1710326	CLDND1	PC.aa.C36.0	0.02(0.00,0.03)	0.03	-0.13(-0.22,-0.03)	0.01	-0.24(-0.46,-0.02)	0.03
cg06235438	ITGAL	ILMN_1710326	CLDND1	PC ae C38:0	0.02(0.00,0.03)	0.03	-0.11(-0.21,-0.02)	0.02	-0.25(-0.48,-0.03)	0.03
cg10825315	TSHR	ILMN_1773650	LRRN3	LPC (18:2)	-0.02(-0.04,0.00)	0.03	-0.09(-0.19,0.01)	0.07	0.12(0.01,0.23)	0.03
cg14753356		ILMN_1773650	LRRN3	LPC (18:2)	0.03(0.00,0.07)	0.03	0.14(-0.01,0.30)	0.07	0.11(0.01,0.22)	0.04
cg22851561	C14orf43	ILMN_1773650	LRRN3	LPC (18:2)	-0.03(-0.06,0.00)	0.03	-0.13(-0.27,0.01)	0.08	0.12(0.00,0.23)	0.05
cg22851561	C14orf43	ILMN_1773650	LRRN3	PC ae C38:0	-0.03(-0.06,0.00)	0.03	-0.15(-0.28,-0.01)	0.03	0.11(0.00,0.22)	0.05
cg24540678		ILMN_1773650	LRRN3	LPC (18:2)	-0.07(-0.13,0.00)	0.02	-0.26(-0.58,0.06)	0.11	0.12(0.01,0.23)	0.03
cg26729380	TNF	ILMN_1773650	LRRN3	PC ae C34:3	-0.06(-0.12,0.00)	0.03	0.33(0.15,0.51)	2.70E-04	-0.12(-0.23,0.00)	0.05
cg27449150		ILMN_1773650	LRRN3	LPC (18:2)	0.04(0.00,0.08)	0.02	-0.34(-0.55,-0.13)	1.70E-03	-0.25(-0.47,-0.03)	0.03
cg27449150		ILMN_1710326	CLDND1	PC ae C38:0	-0.03(-0.07,0.00)	0.04	-0.27(-0.48,-0.06)	0.01	0.12(0.00,0.23)	0.05
cg27449150		ILMN_1773650	LRRN3	PC ae C38:0	-0.04(-0.07,0.00)	0.03	-0.19(-0.39,0.01)	0.07	0.12(0.01,0.23)	0.03

F3										
cg00295485	UXS1	ILMN_1773650	LRRN3	LPC (18:2)	-0.02(-0.16 , 0.12)	0.38	0.20(-0.26 , 0.65)	0.4	-0.66(-1.25 , -0.06)	0.03
cg00893603	ATP8A2	ILMN_1773650	LRRN3	LPC (18:2)	-0.02(-0.12 , 0.07)	0.31	-0.03(-0.33 , 0.28)	0.86	-0.64(-1.24 , -0.04)	0.04
cg01208318		ILMN_1773650	LRRN3	LPC (18:2)	-0.01(-0.11 , 0.08)	0.4	0.01(-0.31 , 0.33)	0.94	-0.65(-1.24 , -0.05)	0.04
cg01731783	C14orf43	ILMN_1710326	CLDND1	PC ae C38:0	-0.05(-0.19 , 0.1)	0.27	0.45(-0.37 , 1.28)	0.29	-0.26(-0.88 , 0.35)	0.4
cg04039799	NAV2	ILMN_1773650	LRRN3	PC ae C38:0	0.02(-0.12 , 0.16)	0.4	0.17(-0.51 , 0.84)	0.63	0.44(-0.14 , 1.02)	0.15
cg04425624	TNF	ILMN_1773650	LRRN3	PC ae C38:0	0.15(-0.17 , 0.47)	0.17	0.13(-1.08 , 1.34)	0.83	0.43(-0.16 , 1.03)	0.16
cg05824218	RARA	ILMN_1710326	CLDND1	PC ae C38:0	0.03(-0.09 , 0.16)	0.31	-0.08(-0.99 , 0.82)	0.86	-0.23(-0.85 , 0.39)	0.48
cg06235438	ITGAL	ILMN_1710326	CLDND1	PC.aa.C36.0	-0.01(-0.05 , 0.04)	0.36	0.12(-0.26 , 0.5)	0.53	-0.14(-0.79 , 0.51)	0.68
cg06235438	ITGAL	ILMN_1710326	CLDND1	PC ae C38:0	-0.02(-0.07 , 0.04)	0.29	0.24(-0.12 , 0.59)	0.19	-0.26(-0.87 , 0.35)	0.4
cg10825315	TSHR	ILMN_1773650	LRRN3	LPC (18:2)	-0.04(-0.16 , 0.09)	0.29	-0.21(-0.63 , 0.2)	0.33	-0.62(-1.22 , -0.03)	0.04
cg14753356		ILMN_1773650	LRRN3	LPC (18:2)	-0.07(-0.27 , 0.13)	0.24	-0.39(-1.04 , 0.25)	0.24	-0.61(-1.2 , -0.02)	0.05
cg22851561	C14orf43	ILMN_1773650	LRRN3	LPC (18:2)	-0.05(-0.2 , 0.1)	0.26	-0.13(-0.61 , 0.35)	0.6	-0.63(-1.23 , -0.03)	0.04
cg22851561	C14orf43	ILMN_1773650	LRRN3	PC ae C38:0	0.03(-0.07 , 0.13)	0.27	0.28(-0.18 , 0.74)	0.24	0.41(-0.17 , 0.99)	0.17
cg24540678		ILMN_1773650	LRRN3	LPC (18:2)	0.00(-0.47 , 0.46)	0.49	-0.33(-1.89 , 1.23)	0.68	-0.65(-1.24 , -0.05)	0.04
cg26729380	TNF	ILMN_1773650	LRRN3	PC.ae.C34.3	0.02(-0.13 , 0.17)	0.41	-0.26(-1.15 , 0.63)	0.57	0.07(-0.53 , 0.67)	0.81
cg27449150		ILMN_1773650	LRRN3	LPC (18:2)	-0.12(-0.38 , 0.14)	0.19	-0.32(-1.16 , 0.52)	0.46	-0.62(-1.21 , -0.02)	0.05
cg27449150		ILMN_1710326	CLDND1	PC ae C38:0	-0.07(-0.24 , 0.11)	0.22	0.79(-0.02 , 1.6)	0.06	-0.32(-0.92 , 0.29)	0.31
cg27449150		ILMN_1773650	LRRN3	PC ae C38:0	0.07(-0.11 , 0.26)	0.22	0.65(-0.15 , 1.45)	0.12	0.38(-0.19 , 0.96)	0.2

Additionally, the methylation and gene expression were exchanged as outcome and mediator in the model to test the potential reverse mediation effects of gene expression for the associations between smoking and DNA methylation. In the KORA F4, 49 gene expressions showed noteworthy mediation effects on the association between methylation and smoking (corrected for multiple testing using Bonferroni criteria, p-value <7E-5, Table S3 in the Appendix). Of the 49 genes, mediation effects of *LRRN3* on the methylation level of cg12593793 and cg21280392 also had a significant mediation effect in the KORA F3 dataset (p-value <0.05, without adjustment of multiple testing). The reversed mediation effects were found for the seven CpG sites that mediated the association between *LRRN3* expression and smoking, suggesting a potential feedback effect of methylation on the genes expression.

3.3 Metabolite markers of incident MI

In this study, a targeted-metabolomics approach was used in prospective cohorts, to identify metabolites related to incident MI. In addition, the association between the metabolites and CRP was investigated to understand the relationship between any identified metabolites and established MI risk factors.

3.3.1 Population Characteristics

Of the 1342 study participants with metabolite profile of fasting serum sample available in the KORA S4, 67 cases of incident MI occurred during a mean follow-up period (±SD) of 5.5 (±2.6) years. Overall, cases of MI occurred more frequently occurred in males than in females, had significantly higher prevalence of Type 2 diabetes and higher levels of CRP: On the other hand these participants exhibit lower levels of HDL-cholesterol and were less physically active than participants without incident MI (Table 15). Within the S2 case-cohort 112 incident cases of MI occurred

based on a mean follow-up period (±SD) of 10 (±3.9) years (Table 15). The study base in the case-cohort of MONICA/KORA S2 was younger with mean (±SD) age (59±7 years in incident MI patients and 53±10 years in non-MI participants) and had a higher percentage of smokers (33.9% in incident MI patients and 23.5% in non-MI participants) in comparison with the KORA S4 study.

Table 15. Population characteristics of discovery and replication dataset

Mean ±standard deviation (SD) is provided when appropriate; p-values are calculated by Mann-Whitney test for continuous data and fisher test for categorical data. [a] ≥20 g/day for women; ≥40 g/day for men; [b] values are presented as geometric mean ±SD. Abbreviation: MI, myocardial infarction.

	Discovery: KORA S4			**Replication: KORA S2**		
	Incident MI	**non-MI**	***P***	**Incident MI**	**non-MI**	***P***
N	67	1275		112	549	
Age	65.5 ±5.2	63.8 ±5.4	0.01	59.71 ±7.53	53.61 ±10.95	5.30E-14
Gender (male,%)	52(77.9%)	623(48.9%)	0.02	85(75.89%)	271(49.36%)	2.40E-07
BMI(kg/m2)	29.6 ±4.8	28.3 ±4.2	0.03	28.28 ±3.87	27.05 ±4.03	3.10E-04
Waist-to-hip ratio	0.95 ±0.07	0.9 ±0.08	1.40E-06	0.93 ±0.07	0.87 ±0.08	6.30E-11
Type 2 diabetes (%)	14(20.9%)	104(8.16%)	1.60E-03	20(17.85%)	18(3.27%)	1.90E-07
Smoking						
Non-smoker (%)	26 (38.2%)	663 (49.7%)	0.08	32 (28.6%)	269 (49.0%)	7.00E-05
Former smoker (%)	30 (44.1%)	463 (36.3%)	0.2	42 (37.5%)	151 (27.5%)	0.04
Smoker (%)	12 (17.7%)	117 (13.9%)	0.37	38 (33.9%)	129 (23.5%)	0.02
Alcohol intake(%)[a]	12 (17.7%)	262 (20.6%)	0.65	28 (25%)	153 (27.86%)	0.56
Physical activity (% >1 h per week)	16 (23.5%)	560 (43.9%)	9.50E-04	30 (26.78%)	208 (37.7%)	0.03
Diastolic blood pressure (mm Hg)	81.2 ±11.9	80.1 ±10.2	0.37	80.81 ±11.63	80.13 ±10.97	0.9
Systolic blood pressure (mm Hg)	142.9 ±24.3	135.2 ±19.8	0.01	143.77 ±22.07	133.35 ±18.35	5.80E-07
total Cholesterol (mg/dl)	240.4 ±38.7	244.1 ±41.2	0.32	260.7 ±50.44	240.34 ±43.19	1.90E-05
HDL-Cholesterol (mg/dl)	53.1 ±16.8	58.9 ±16.5	2.00E-03	50.81 ±15.51	58.26 ±16.67	9.20E-07
LDL-Cholesterol (mg/dl)	157.0 ±35.4	154.2 ±39.3	0.55	171.9 ±48.74	149.75 ±40.74	7.90E-08
Total/HDL cholesterol	4.91 ±1.6	4.43 ±1.33	0.01	5.6 ±2.21	4.5 ±1.9	6.90E-12
CRP (mg/L)[b]	1.69 ±6.63	2.87 ±5.15	4.50E-04	1.57 ±7.69	2.99 ±6.42	2.70E-08
Statin user (%)	4 (5.97%)	109(8.54%)	0.65	0 (0%)	3 (0.54%)	1

3.3.2 Identification of candidate biomarkers of incident MI

The candidate biomarkers were discovered in the KORA S4 study and replicated in the MONICA/KORA S2 cohort. The analysis procedure is illustrated in Figure 14.

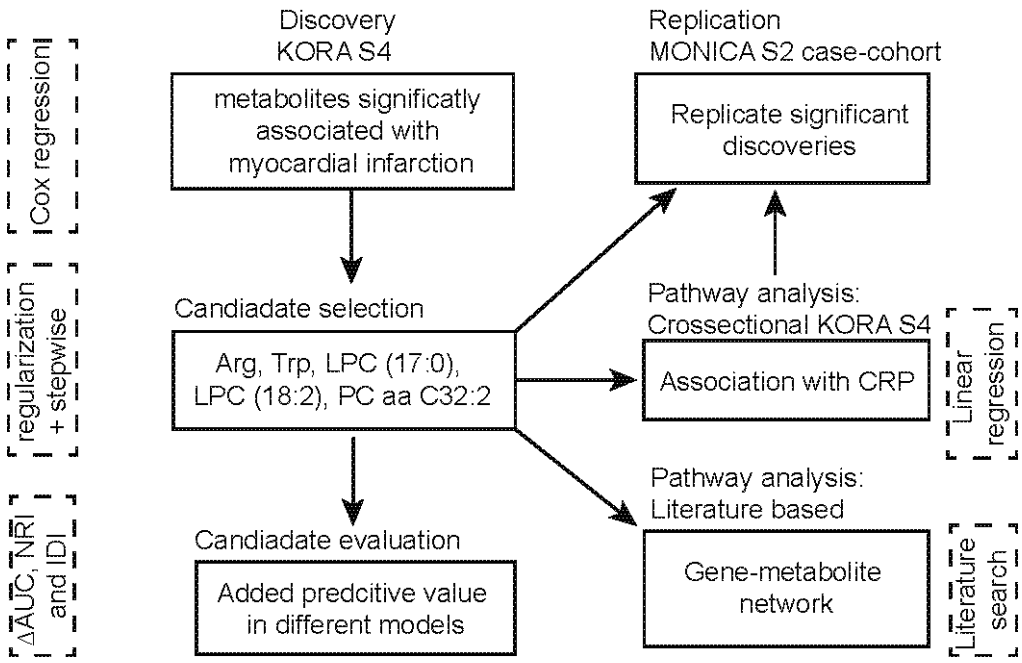

Figure 14. Flow diagram illustrating the analysis strategy for the study of metabolite markers of incident MI

Abbreviations: Arg, arginine; Trp, tryptophan; LPC, lysophosphatidylcholine; CRP, C-reactive protein

Cox regression analysis, using the discovery KORA S4 dataset, identified 21 metabolites which are significantly associated with incident MI in both the crude model and multivariate model. After further adjustment for CRP, 16 of these metabolites remained significant (p-value <0.05). The 16 metabolites belong to three groups: amino acids, lysophosphatidylcholines, phosphatidylcholines. These metabolites displayed high within-group correlations and high correlations with cardiovascular disease risk factors, e.g. BMI, LDL and HDL (Figure 15).

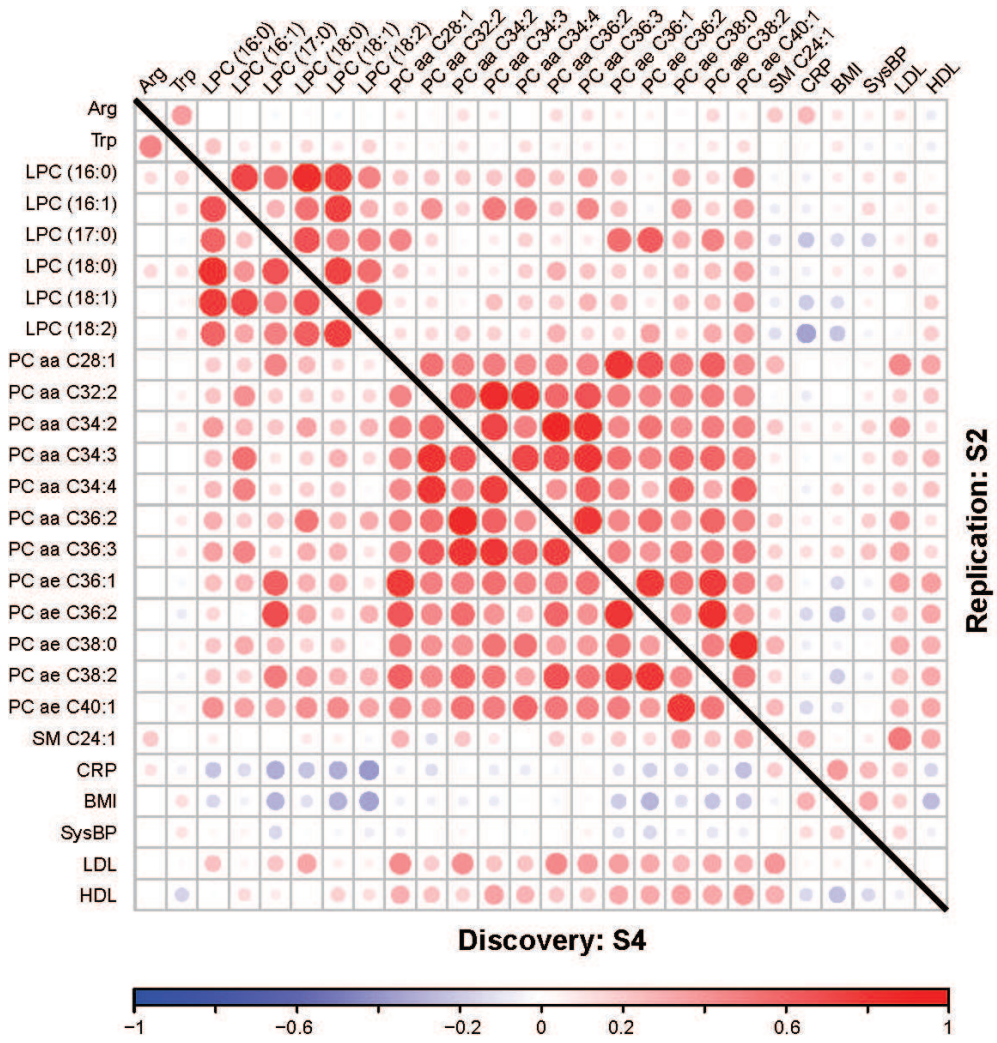

Figure 15. Correlations of the metabolites and risk factors of MI

Abbreviations: Arg, arginine; Trp, tryptophan; LPC, lysophosphatidylcholine; PC, phosphatidylcholine; aa, diacyl; ae, acyl-akyl; CRP, C-reactive protein; BMI, body mass index; SysBP, systolic blood pressure; LDL, low density lipoprotein; HDL, high density lipoprotein

To find an independent predictor, regularization followed by a step-wise selection analysis was applied to the 16 metabolites. Five metabolites (i.e. arginine, tryptophan, LPC (17:0), LPC (18:2) and PC aa C32:2) were selected as candidate biomarkers (Table 16). The risk of incident MI increases by 30% to 40% for every SD increase of the log transformed arginine level, whilst the risk is reduced by 20% to 40% for tryptophan, LPC (17:0) , LPC (18:2) and PC aa C32:2 (Table 16). In the replication study, arginine, LPC (17:0) and LPC (18:2) were found to be greatly associated with

MI in both crude and multivariate models with similar effective sizes as in the discovery study; while the associations of tryptophan and PC aa C32:2 could not be replicated. A meta-analysis of the associations showed consistent and significant associations between arginine, LPC (17:0) and LPC (18:2) and incident MI (Table 16).

Table 16. Associations of selected metabolites with incident MI in the discovery (KORA S4) and replication (KORA S2) datasets

Crude model adjusted for age and sex; multivariate model adjusted for BMI, diabetes, systolic blood pressure, smoking status, (current smoker, former smoker and never smoker), alcohol consumption(≥20 g/day for women; ≥40 g/day for men), total cholesterol, HDL cholesterol in addition to crude model; CRP was further adjusted in the final model. Values in the table are provided as estimated hazard ratio (95% confidence interval) for each SD increase of the log-transformed metabolite concentration. Abbreviation: MI, myocardial infarction; HR, hazardous ratio; CI: confidence interval.

	S4		**S2**		**Meta-analysis**		**Heterogeneity**	
	HR (95% CI)	***P***	**HR (95% CI)**	***P***	**HR(95%CI)**	***P***	**Q**	***P***
Crude model								
Arg	1.40(1.10,1.78)	0.006	1.26(1.05,1.52)	0.01	1.31(1.14,1.49)	1.7E-04	0.51	0.47
Trp	0.79(0.62,1.00)	0.05	1.20(0.99,1.45)	0.06	1.03(0.89,1.19)	0.71	7.4	0.01
LPC (17:0)	0.65(0.52,0.82)	2.5E-04	0.77(0.64,0.93)	0.01	0.74(0.64,0.84)	1.8E-05	1.69	0.19
LPC (18:2)	0.71(0.56,0.91)	5.3e-5	0.67(0.56,0.82)	4.8E-05	0.64(0.56,0.75)	3.80E-09	0.6	0.44
PC aa C32:2	0.60(0.46,0.77)	2.1E-04	1.09(0.91,1.33)	0.32	0.85(0.73,0.98)	0.03	8.78	3.0E-03
Multivariate Model								
Arg	1.35(1.06,1.72)	0.016	1.23(1.02,1.47)	0.027	1.27(1.10,1.46)	8.9E-04	0.39	0.53
Trp	0.72(0.57,0.92)	0.01	1.45(1.18,1.79)	4.0E-04	1.03(0.52,2.03)	0.93	19.14	1.2E-05
LPC (17:0)	0.69(0.53,0.89)	0.005	0.72(0.57,0.90)	0.004	0.72(0.61,0.85)	1.0E-04	0.21	0.65
LPC (18:2)	0.65(0.49,0.87)	0.004	0.79(0.64,0.96)	0.019	0.74(0.62,0.88)	8.70E-04	1.18	0.28
PC aa C32:2	0.64(0.50,0.81)	2.2E-04	0.95(0.77,1.17)	0.627	0.80(0.51,1.24)	0.32	7.99	4.7E-03
Multivariate Model + CRP								
Arg	1.32(1.04,1.69)	0.024	1.17(0.97,1.42)	0.096	1.23(1.07,1.42)	4.7E-03	0.56	0.46
Trp	0.77(0.60,0.99)	0.044	1.42(1.16,1.75)	7.4E-04	1.05(0.58,1.93)	0.86	14.44	0
LPC (17:0)	0.71(0.55,0.93)	0.01	0.75(0.60,0.95)	0.014	0.75(0.64,0.89)	7.2E-04	0.28	0.6
LPC (18:2)	0.7(0.52,0.94)	0.017	0.84(0.68,1.03)	0.08	0.79(0.67,0.93)	0.01	1.03	0.31
PC aa C32:2	0.67(0.52,0.86)	0.002	1.04(0.84,1.30)	0.71	0.52	0.48	3.6E-03	0.52

Besides the five selected candidates, four additional lyso-PCs (LPC (16:0), LPC (16:1), LPC (18:0) and LPC (18:2)) showed consistent negative association with MI in both datasets (Figure 16).

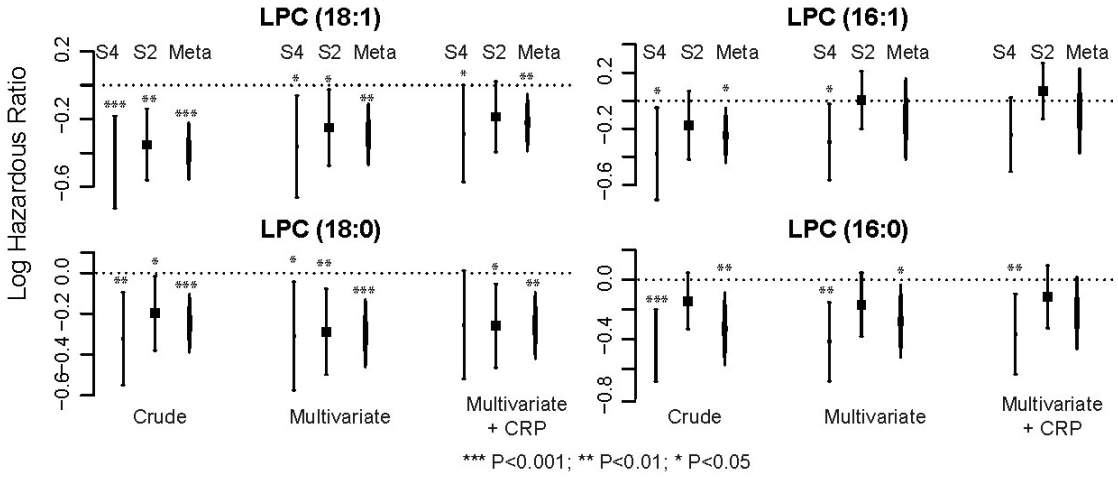

Figure 16. Associations of incident MI with other LPCs in the discovery (KORA S4) and replication (KORA S2) datasets

Associations between incident MI with four lyso-PCs (LPC (18:1), LPC (18:0), LPC (16:1), LPC (16:0)) were calculated in cox regression models. Crude model adjusted for age and sex; multivariate model adjusted for BMI, diabetes, systolic blood pressure, smoking status (current smoker, former smoker and never smoker), alcohol consumption (≥20 g/day for women; ≥40 g/day for men), total cholesterol, HDL cholesterol in addition to crude model; CRP was further adjusted in the last model. Random effect model was used for meta-analysis. Associations were shown as log hazards ratios.

In the discovery dataset, the addition of the five metabolites to the crude and multivariate models significantly increased C statistics, net reclassification improvement and integrated discrimination improvement (Table 17). After adding the five candidate metabolites to the Framingham score, the C statistics increased by 0.8 (P=0.001), yielding a categorical net reclassification improvement of 0.21 (P <0.0001, categories: 0-10%, 11-20% and > 20%) and the integrated discrimination index (IDI) of 0.15 (P <0.0001). Improvement of prediction of the five metabolites in the replication dataset was not significant (Table 17).

Table 17. Added value of selected metabolites in prediction of incident MI in discovery the KORA S4 and S2

Values in the table are provided as Mean (95% CI). Crude model is adjusted for age, sex; multivariate model is adjusted for age, sex, BMI, diabetes, systolic blood pressure, smoking, alcohol consumption, total cholesterol, HDL cholesterol, CRP. Abbreviations: AUC, area under the curve; NRI, net reclassification index; IDI, integrated discrimination improvement.

	AUC	***P***	**NRI categorical** [a]	***P***	**IDI**	***P***
S4						
Crude model	0.67(0.60,0.74)					
+selected metabolites	0.77(0.73,0.83)	8.90E-05	0.19(0.06,0.32)	0.004	0.11(0.06,0.15)	<2E-16
Multivariate model	0.68(0.62,0.74)					
+ selected metabolites	0.76(0.72,0.82)	6.60E-05	0.17(0.04,0.30)	0.01	0.10(0.06, 0.14)	<2E-16
Multivariate model + CPR	0.69(0.63,0.76)					
+ selected metabolites	0.76(0.71, 0.82)	0.002	0.12(0.00, 0.25)	0.04	0.08(0.04,0.11)	1.00E-05
Framingham score	0.70(0.64,0.76)					
+ selected metabolites	0.78(0.73,0.84)	0.001	0.22(0.09,0.36)	9.80E-04	0.15(0.10,0.20)	<2E-16
S2						
Crude model	0.70(0.65,0.75)					
+ selected metabolites	0.72(0.67,0.77)	0.12	0.04(-0.05,0.13)	0.42	0.02(0.007,0.04)	0.005
Multivariate model	0.76(0.72,0.81)					
+ selected metabolites	0.77(0.73,0.82)	0.33	0.07(-0.04,0.18)	0.2	0.02(0.003, 0.04)	0.02
Multivariate model + CRP	0.77(0.72,0.81)					
+ selected metabolites	0.77(0.73, 0.82)	0.48	0.06(-0.03, 0.16)	0.2	0.02(0.005,0.04)	0.01
Framingham score	0.74(0.70,0.79)					
+ selected metabolites	0.77(0.72,0.80)	0.14	0.096(-0.005,0.20)	0.06	0.008(-0.005,0.02)	0.21

[a] categories were set at 0~10%, 11~20% and over 21%

3.3.3 The five candidate metabolites were strongly associated with CRP

It is noteworthy that in the established prediction model with the five candidate biomarkers, the estimates of CRP were reduced and no longer statistical significant (Figure 17, p-value = 0.74 in the discovery, p-value = 0.23 in the replication, p-value = 0.22 in meta-analysis). With this said, no changes in other covariates were observed (Table 18). To elaborate, age had a hazard ratio (95% CI) of 1.27 (0.98, 1.65) before and a ratio of 1.24 (0.94, 1.63) after adding the five candidate biomarkers in the discovery data set; 2.64 (1.99, 3.5) and 2.67 (2.00, 3.57) in the replication dataset (Table 18).

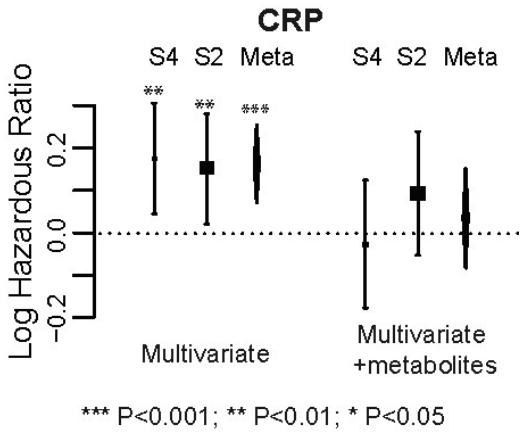

Figure 17. Associations of incident MI with CRP in the discovery (KORA S4) and replication (KORA S2) datasets

The association between CRP and MI was attenuated after adding the selected metabolites in the cox regression model. Associations are shown as log hazards ratios. Random effect meta-analysis was performed

Table 18. Change of CRP association with incident MI after adding selected metabolites

Estimates of the variables in the cox regression model were shown in the tables.

	S4				**S2**				**Meta-analysis**			
	HR (95% CI)	P	**HR (95% CI)**	P	**HR (95% CI)**	P	**HR (95% CI)**	P	**HR (95% CI)**	**P**	**HR**	**P**
Arginine			1.69 (1.26,2.27)	4.4E-04			1.08 (0.87,1.32)	0.49			1.33 (0.85,2.08)	0.21
Tryptophan			0.75 (0.55,1.01)	0.06			1.47 (1.18,1.84)	7.4E-04			0.99 (0.45,2.17)	0.97
LPC(17:0)			0.66 (0.50,0.87)	3.2E-03			0.71 (0.54,0.94)	0.02			0.73 (0.59,0.89)	2.5E-03
LPC (18:2)			0.92 (0.65,1.29)	0.62			0.88 (0.68,1.13)	0.30			0.89 (0.73,1.09)	0.26
PC aa C32:2			0.78 (0.60,1.02)	0.07			0.98 (0.78,1.24)	0.88			0.89 (0.71,1.11)	0.29
Age	1.27 (0.98,1.65)	0.08	1.24 (0.94,1.63)	0.13	2.64 (1.99,3.5)	1.23E-11	2.67 (2.00,3.57)	3.0E-11	1.83 (0.89,3.75)	0.1	1.82 (0.85,3.86)	0.12
Sex[a]	0.28 (0.14,0.55)	2.61E-04	0.26 (0.12,0.54)	3.7E-04	0.33 (0.19,0.56)	3.82E-05	0.36 (0.21,0.64)	4.7E-04	0.31 (0.2,0.47)	4.01E-08	0.32 (0.20,0.50)	7.8E-07
BMI	1.23 (0.95,1.59)	0.12	1.19 (0.90,1.58)	0.22	0.88 (0.7,1.12)	0.31	0.79 (0.62,1.02)	0.07	1.04 (0.75,1.43)	0.82	0.97 (0.65,1.44)	0.87
Diabetes	1.8 (0.92,3.54)	0.09	1.85 (0.93,3.66)	0.08	3.52 (2.02,6.13)	9.21E-06	4.91 (2.74,8.79)	8.4E-08	2.59 (1.35,4.97)	4.20E-03	3.07 (1.18,7.98)	0.02
systolic blood pressure	1.25 (0.97,1.6)	0.08	1.19 (0.92,1.54)	0.18	1.37 (1.12,1.66)	1.69E-03	1.39 (1.14,1.70)	9.4E-04	1.32 (1.13,1.54)	3.83E-04	1.32 (1.13,1.54)	5.7E-04
Current smoker [b]	0.87 (0.48,1.59)	0.66	0.80 (0.44,1.46)	0.47	1.59 (0.95,2.66)	0.08	1.70 (0.99,2.91)	0.05	1.2 (0.67,2.15)	0.54	1.18 (0.56,2.48)	0.66
Former smoker [b]	1.54 (0.73,3.24)	0.25	1.22 (0.57,2.62)	0.61	2.27 (1.33,3.9)	2.81E-03	2.18 (1.23,3.84)	0.01	1.99 (1.29,3.08)	2.00E-03	1.73 (0.99,3.02)	0.05
Alcohol intake[c]	0.61 (0.3,1.22)	0.16	0.60 (0.29,1.24)	0.17	0.79 (0.5,1.27)	0.33	0.55 (0.33,0.92)	0.02	0.73 (0.5,1.08)	0.11	0.57 (0.37,0.87)	0.01
Total cholesterol	1.01 (0.77,1.31)	0.97	1.12 (0.85,1.47)	0.42	1.52 (1.27,1.82)	4.71E-06	1.54 (1.27,1.86)	5.4E-06	1.25 (0.83,1.88)	0.28	1.34 (0.98,1.83)	0.07
HDL cholesterol	1.01 (0.74,1.38)	0.95	0.97 (0.71,1.34)	0.87	0.73 (0.57,0.93)	0.01	0.79 (0.61,1.02)	0.07	0.84 (0.61,1.17)	0.31	0.86 (0.70,1.05)	0.14
CRP	1.19 (1.04,1.35)	0.01	1.05 (0.80,1.38)	0.74	1.16 (1.02,1.32)	0.02	1.10 (0.94,1.27)	0.23	1.17 (1.07,1.28)	5.58E-04	1.08 (0.69,1.67)	0.73

[a] women compared with men; [b] compared with never smokers; [c] ≥20 g/day for women; ≥40 g/day for men

The levels of CRP had significant association with the arginine, LPC (17:0), LPC (18:0), and PC aa C32:2 (Table 19) in both the discover and replication datasets. Tryptophan was significantly associated with CRP in the discovery dataset but not in the replication dataset. The combination of the potential biomarkers additionally explains for the approximate 10 % variation of CRP levels, from 14% to 24% in the discovery dataset and from 22% to 32% in the replication dataset (Table 19). Likelihood ratio test between the regression model of CRP with and without the metabolites also show significant improvement in goodness of fit ($P < 0.001$ in both KORA S4 and KORA S2) after adding the metabolites to the model (Table 19).

Table 19. Associations between CRP and selected metabolites in the discovery (KORA S4) and replication (KORA S2) dataset

The β estimates for the metabolites indicate the corresponding change in the log-transformed concentration of CRP as one SD change in the log-transformed metabolite concentrations (ratio). [a] Change of R^2 in regression model after adding the selected metabolites in the model with age, sex, bmi, diabetes, systolic blood pressure, smoking status(current smoker, former smoker and never smoker), alcohol consumption(≥20 g/day for women; ≥40 g/day for men), total cholesterol, HDL cholesterol. [b] likelihood ratio test was used to compare the model with and without the selected metabolites. Abbreviations: CI, confidence interval

	S4		**S2**	
Metabolites	**β (95% CI)**	***P***	**β (95% CI)**	***P***
Arginine	0.07(0.01,0.12)	0.02	0.24(0.16,0.32)	2.8E-09
Tryptophan	-0.11(-0.16,-0.05)	2.1E-04	0.04(-0.04,0.13)	0.3
LPC (17:0)	-0.27(-0.33,-0.21)	8.8E-19	-0.19(-0.28,-0.11)	2.2E-05
LPC (18:2)	-0.31(-0.37,-0.25)	1.5E-24	-0.28(-0.37,0-0.20)	2.1E-11
PC aa C32:2	-0.15(-0.21,-0.09)	7.8E-07	-0.17(-0.26,-0.08)	1.3E-04
[a] Increase of R^2	0.104		0.101	
[b]Likelihood ratio test	$P < 2.2E-16$		$P < 2.2E-16$	

3.3.4 Sensitivity analysis

Sensitivity analyses were used to test whether the associations between the five metabolites and incident MI were confounded by statins and diabetes. The associations of metabolites with incident MI remained significant after excluding statin users (N = 68 in the discovery dataset; N = 3 in the replication dataset) (Table

20). The effective size of the associations remained unaffected after exclusion of diabetic patients (N = 118 in the discovery dataset; N = 38 in the replication dataset), which is the strongest risk factor in the model (Table 20).

Furthermore, the influence of fasting status was investigated in the discovery data set (KORA S4) by including non-fasting samples in the association analysis. In total, data from 203 non-fasting serum samples were added to the analysis, including 22 MI cases and 181 non-cases. Arginine, LPC (17:0), LPC (18:2), and PC aa C32:2 were still significantly associated with incident MI in the multivariate model, but, the relationship between tryptophan and MI were not significant in any of the three models (Table 20).

Table 20. Sensitivity analysis of associations between selected metabolites and incident MI

Values in the table are provided as estimated hazard ratio (95% confidence interval) for each standard deviation increase of the log-transformed metabolite concentrations (ratio). Abbreviation: MI, myocardial infarction; HR, hazardous ratio; CI: confidence interval

		Arginine	Tryptophan	LPC (17:0)	LPC (18:2)	PC aa C32:2
Exclude participants with diabetes						
S4 (54/1171)	HR	1.31	0.77	0.69	0.67	0.68
	(95% CI)	(1.02,1.67)	(0.60,0.99)	(0.53,0.90)	(0.49,0.90)	(0.53,0.87)
	P	0.03	0.04	0.006	0.008	2.40E-03
S2 (92/532)	HR	1.13	1.28	0.71	0.72	0.81
	(95% CI)	(0.93,1.38)	(1.02,1.62)	(0.56,0.90)	(0.57, 0.90)	(0.63,1.04)
	P	0.22	0.04	0.005	0.005	0.1
Exclude statin users						
S4 (64/1161)	HR	1.30	0.76	0.69	0.70	0.64
	(95% CI)	(1.01,1.70)	(0.58,0.98)	(0.52,0.91)	(0.50,0.97)	(0.49, 0.83)
	P	0.04	0.04	9.50E-03	0.03	7.80E-04
S2 (112/546)	HR	1.17	1.42	0.75	0.83	1.04
	(95% CI)	(0.98,1.42)	(1.15,1.74)	(0.60,0.95)	(0.68, 1.03)	(0.84,1.30)
	P	0.08	7.40E-04	0.015	0.09	0.71
Include non-fasting samples						
S4 (86/1417)	HR	1.31	0.87	0.79	0.75	0.71
	(95% CI)	(1.06,1.62)	(0.70,1.08)	(0.63,0.99)	(0.60,0.94)	(0.57,0.89)
	P	0.014	0.198	0.04	0.011	0.002

Owing to the fact that there was a small number of cases in this study, in subgroups with different ages (>60 yeas v.s. ≤60 years), sex (male and female), or smoking habits (never smoker, former smoker and current smoker), the associations between the three selected metabolites and incident MI were not significant in neither the discovery nor replication datasets. However, heterogeneity analysis showed no significant differences for the associations in subgroups with different age, gender or smoking habits (Table 21).

Table 21. Subgroup analysis of the MI-associated metabolites

Association between the candidate metabolite and MI were assess in different age, gender and smoking subgroup using cox-regression models with adjustment of age, sex, BMI, alcohol consumption, systolic blood pressure, diabetes status, total cholesterol, HDL cholesterol and CRP.

	Arginine	**Tryptophan**	**LPC(17:0)**	**LPC(18:2)**	**PC aa C32:2**
S4					
Age					
<=60 years					
HR (95% CI)	1.37(1.03,1.80)	0.84(0.64,1.09)	0.78(0.57,1.07)	0.92(0.67,1.27)	0.63(0.48,0.82)
P	0.03	0.19	0.13	0.61	0.001
>60 years					
HR (95% CI)	1.31(0.69,2.49)	0.78(0.37,1.62)	0.39(0.18,0.83)	0.98(0.47,2.04)	0.58(0.29,1.16)
P	0.41	0.50	0.02	0.95	0.121
Heterogeneity	0.05	0.02	2.78	0.03	0.01
P	0.83	0.88	0.10	0.85	0.90
Sex					
Male					
HR (95% CI)	1.44(1.08,1.92)	0.89(0.66,1.19)	0.79(0.57,1.08)	0.73(0.52,1.04)	0.72(0.54,0.96)
P	0.01	0.42	0.14	0.08	0.03
Female					
HR (95% CI)	1.11(0.63,1.95)	0.49(0.28,0.85)	0.50(0.26,0.94)	0.93(0.47,1.84)	0.49(0.28,0.88)
P	0.73	0.01	0.03	0.84	0.02
Heterogeneity	1.35	0.38	1.60	3.49	0.67
P	0.24	0.54	0.21	0.06	0.41
Smoking					
Never smoker					
HR (95% CI)	1.44(0.78,2.67)	0.85(0.45,1.57)	0.67(0.33,1.35)	0.81(0.40,1.66)	0.48(0.24,0.98)
P	0.25	0.60	0.26	0.57	0.04
Former smoker					
HR (95% CI)	1.42(0.99,2.04)	0.86(0.58,1.27)	0.77(0.50,1.17)	0.69(0.43,1.08)	0.57(0.39,0.81)
P	0.06	0.45	0.22	0.10	0.002
Heterogeneity	0.15	0.15	0.10	0.002	0.001
P	0.69	0.70	0.75	0.96	0.98
Current smoker					
HR (95% CI)	1.19(0.76,1.88)	0.66(0.42,1.04)	0.62(0.38,1.01)	0.81(0.48,1.35)	0.91(0.58,1.41)
P	0.45	0.07	0.05	0.42	0.67
Heterogeneity	2.18	7.4E-05	0.03	0.41	0.23
P	0.14	0.99	0.86	0.52	0.63

S2					
Age					
<=60 years					
HR (95% CI)	1.13(0.88,1.44)	1.42(1.09,1.86)	0.67(0.51,0.89)	0.69(0.52,0.92)	0.82(0.59,1.15)
P	0.34	0.01	0.01	0.01	0.25
>60 years					
HR (95% CI)	1.48(1.03,2.13)	1.45(1.02,2.05)	0.66(0.45,0.98)	0.68(0.51,0.92)	1.32(0.95,1.82)
P	0.03	0.04	0.04	0.01	0.10
Heterogeneity	3.94	0.00	0.01	0.01	1.52
P	0.05	0.96	0.94	0.93	0.22
Sex					
Male					
HR (95% CI)	1.12(0.90,1.38)	1.36(1.07,1.73)	0.67(0.52,0.86)	0.75(0.60,0.94)	0.99(0.77,1.28)
P	0.32	0.01	0.002	0.01	0.94
Female					
HR (95% CI)	1.30(0.82,2.07)	1.47(0.93,2.32)	1.04(0.63,1.72)	0.84(0.53,1.33)	0.88(0.56,1.40)
P	0.27	0.10	0.87	0.45	0.60
Heterogeneity	0.18	0.16	2.38	0.09	0.34
P	0.67	0.69	0.12	0.77	0.56
Smoking					
Never smoker					
HR (95% CI)	1.16(0.83,1.63)	1.06(0.74,1.52)	0.79(0.49,1.27)	0.99(0.65,1.51)	1.69(0.99,2.89)
P	0.39	0.76	0.33	0.97	0.06
Former smoker					
HR (95% CI)	1.25(0.93,1.68)	1.61(1.13,2.29)	0.58(0.40,0.82)	0.70(0.50,0.98)	0.78(0.54,1.13)
P	0.14	0.01	0.002	0.04	0.19
Heterogeneity	5.37	1.59	1.08	2.65	0.09
P	0.02	0.21	0.30	0.10	0.76
Current smoker					
HR (95% CI)	1.11(0.73,1.70)	1.87(1.22,2.88)	0.72(0.46,1.14)	0.60(0.40,0.91)	1.01(0.66,1.54)
P	0.62	0.004	0.17	0.02	0.96
Heterogeneity	2.18	2.71	0.07	3.96	0.03
P	0.14	0.10	0.79	0.05	0.87

3.3.5 Candidate metabolites associated with CVD related genes

To find further support of the functionality of the five selected metabolites, their association with coronary artery disease related genes reported in GWAS (Deloukas *et al*, 2012) were analyzed by searching a protein-metabolite network. In total, 52 genes from 46 coronary artery disease related loci with genome-wide significance and 123 genes from 104 loci at a 5% false discovery rate (FDR) criteria were evaluated. Among these genes, 12 genes with genome-wide significance (PPAP2B, APOA1, APOA2, APOA5, APOB, APOE, FLT, LPA, PEMT, EDNRA, GUCY1A3, RASD1) and 4 FDR-significant genes (NGF, SCARB1, APOC2, CDKN1A) were linked with

the five metabolites with at most one intermediate enzyme in between (Figure 18). All links were manually checked for biochemical relevance (Table 22).

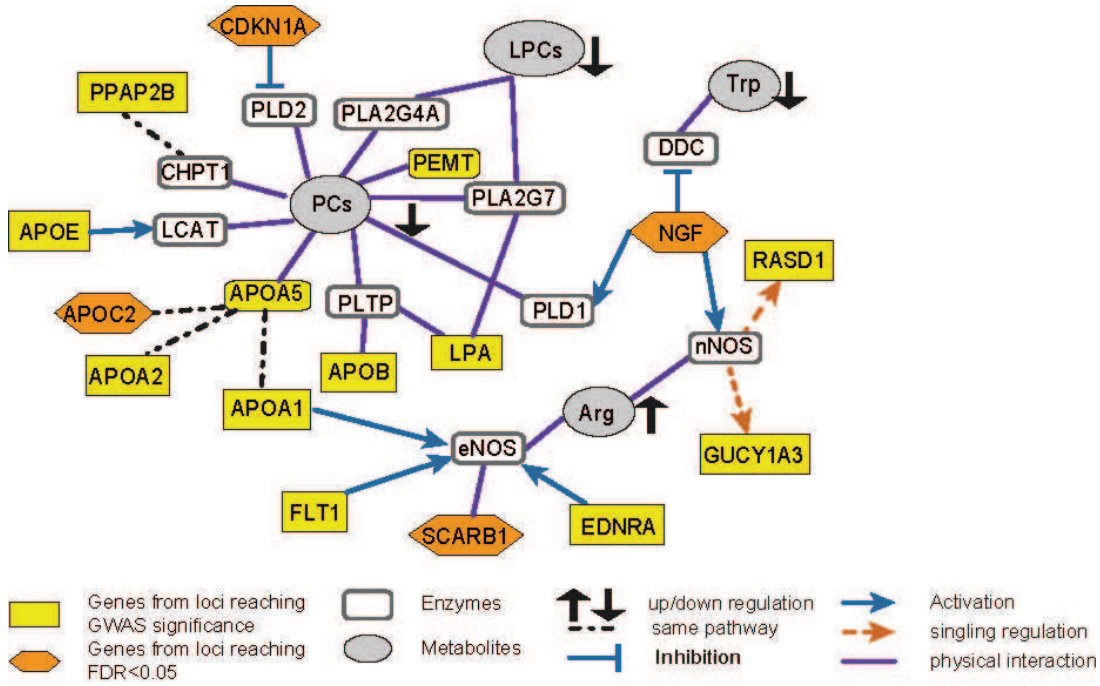

Figure 18. Network and pathway analysis of the selected metabolites
Associations between CAD related genes from GWAS studies and the candidate metabolites.

Table 22. Association between CVD genes and the candidate metabolites

Gene1	Gene2	Interaction
APOA5	APOA1	In the same KEGG pathway PPAR signaling pathway
APOA5	APOA2	Similar function; In the same KEGG pathway PPAR signaling pathway
APOA5	APOC2	chylomicron => TG-depleted chylomicron + long-chain fatty acids + diacylglycerols
Arg	PADI4	L-arginine + H2O = protein L-citrulline + NH3 [RN:R02621]
Arg	RARS	ATP + L-arginine + tRNAArg = AMP + diphosphate + L-arginyl-tRNAArg [RN:R03646]
Arg	NOS3	L-arginine + NADPH + H^+ + O_2 = citrulline + nitric oxide + $NADP^+$ [RN:R00557]
Arg	NOS1	L-arginine + NADPH + H^+ + O_2 = citrulline + nitric oxide + $NADP^+$ [RN:R00557]
CHPT1	PPAP2B	PPAP2B: a 1,2-diacylglycerol 3-phosphate + H_2O = a 1,2-diacyl-sn-glycerol + phosphate [RN:R02239] CHPT1: CDP-choline + 1,2-diacyl-sn-glycerol = CMP + a phosphatidylcholine [RN:R01321]
DDC	NGF	Inhibition: NGF --| DDC [1]
LCAT	APOE	Activation: APOE→LCAT[2]
LPCs	PLA2G4A	(1) 2-lysophosphatidylcholine + H_2O = glycerophosphocholine + a carboxylate [RN:R07291] (2) Phosphatidylcholine + H_2O = 1-acylglycerophosphocholine + a carboxylate
LPCs	PLA2G7	1-alkyl-2-acetyl-sn-glycero-3-phosphocholine + H2O = 1-alkyl-sn-glycero-3-phosphocholine + acetate [RN:R04452]
LPCs	LYPLA1	2-lysophosphatidylcholine + H2O=glycerophosphocholine + a carboxylate;
NOS1	NGF	Activation: NGF→NOS1 [3]
NOS1	RASD1	Binding: RASD1 is activated by NO donors [4, 5]
NOS1	GUCY1A3	NOS=>NO→GUCY1A3: NO activates GUCY1A3 [6]
NOS3	FLT1	FLT1=>VEGF →NOS3 [7]; KEGG: (1) VEGF signaling pathway; (2) PI3K-AKT signaling pathway
NOS3	APOA1	Binding, mediate(APOA1→NOS3) [8, 9]
NOS3	EDNRA	NOS3 induced by vasoconstrictor peptide endothelin-1 (ET-1) via EDNRA [10]
NOS3	SCARB1	Binding: HDL stimulates by SCARB1 binding to NOS3 [8, 11]

PC aa C32:2	PLA2G4A	(1) 2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate [RN:R07291] (2) Phosphatidylcholine + H(2)O = 1-acylglycerophosphocholine + a carboxylate
PC aa C32:2	PEMT	S-adenosyl-L-methionine + phosphatidylethanolamine = S-adenosyl-L-homocysteine + phosphatidyl-N-methylethanolamine
PC aa C32:2	APOA5	
PC aa C32:2	LCAT	phosphatidylcholine + a sterol = 1-acylglycerophosphocholine + a sterol ester [RN:R02114]
PC aa C32:2	PLD2	a phosphatidylcholine + H2O = choline + a phosphatidate [RN:R01310]
PC aa C32:2	CHPT1	CDP-choline + 1,2-diacyl-sn-glycerol = CMP + a phosphatidylcholine [RN:R01321]
PC aa C32:2	PLTP	transfer protein
PC aa C32:2	PLD1	a phosphatidylcholine + H2O = choline + a phosphatidate [RN:R01310]
PC aa C32:2	ABCB4	Transporter: ATP + H2O + xenobioticin = ADP + phosphate + xenobioticout [13]
PC aa C32:2	APOB	Phosphatidylcholine increased apoB synthesis and secretion without affecting the synthesis or secretion of apoA-I [14, 15]
PLD1	NGF	Activation: NGFàPLD1[16]
PLD2	CDKN1A	Inhibition: Cdkn1a ---| PLD2 [17]
PLTP	APOB	transfer ApoB [18]
PLTP	SCARB1	all related to HDL metabolism [19]
PLTP	LPA	transfer LPA [20]
Trp	DDC	(1) 3,4-dihydroxy-L-phenylalanine = dopamine + CO2 [RN:R02080] (2) 5-hydroxy-L-tryptophan = 5-hydroxytryptamine + CO2 [RN:R02701]
LPA	PLA2G7	Positive association [21]

[1]. Li XM, Qi J, Juorio AV, Boulton AA: Reciprocal regulation of the content of aromatic L-amino acid decarboxylase and tyrosine hydroxylase mRNA by NGF in PC12 cells. J. Neurosci. Res. 1997, 47:449–454.
[2]. Zhao, Y., F. E. Thorngate, K. H. Weisgraber, D. L. Williams, and J. S. Parks. 2005. Apolipoprotein E is the major physiological activator of lecithin-cholesterol acyltransferase (LCAT) on apolipoprotein B lipoproteins. Biochemistry. 44: 1013–1025
[3]. Kalisch BE, Demeris CS, Ishak M, Rylett RJ: Modulation of nerve growth factor-induced activation of MAP kinase in PC12 cells by inhibitors of nitric oxide synthase. J. Neurochem. 2003, 87:1321–1332.
[4]. Fang M, Jaffrey SR, Sawa A, Ye K, Luo X, Snyder SH: Dexras1: a G protein specifically coupled to neuronal nitric oxide synthase via CAPON. Neuron 2000, 28:183–193.

[5]. Li X, Cheng C, Fei M, Gao S, Niu S, Chen M, Liu Y, Guo Z, Wang H, Zhao J, Yu X, Shen A: Spatiotemporal expression of Dexras1 after spinal cord transection in rats. Cell. Mol. Neurobiol. 2008, 28:371–388.
[6]. Español AJ, Sales ME: Participation of nitric oxide synthase and cyclo-oxygenase in the signal transduction pathway of ileal muscarinic acetylcholine receptors. Pharmacol. Res. 2000, 42:489–493.
[7]. Duval M, Le Boeuf F, Huot J, Gratton J-P: Src-mediated phosphorylation of Hsp90 in response to vascular endothelial growth factor (VEGF) is required for VEGF receptor-2 signaling to endothelial NO synthase. Mol. Biol. Cell 2007, 18:4659–4668.
[8]. McAllister RM, Morris DM, Weimer CM, Laughlin MH: Effects of high-density lipoprotein on endothelium-dependent vasorelaxation. Appl Physiol Nutr Metab 2010, 35:319–327.
[9]. Mineo C, Shaul PW: HDL stimulation of endothelial nitric oxide synthase: a novel mechanism of HDL action. Trends Cardiovasc. Med. 2003, 13
[10]. Marsen TA, Egink G, Suckau G, Baldamus CA: Tyrosine-kinase-dependent regulation of the nitric oxide synthase gene by endothelin-1 in human endothelial cells. Pflugers Arch. 1999, 438:538–544.
[11]. Gong M, Wilson M, Kelly T, Su W, Dressman J, Kincer J, Matveev SV, Guo L, Guerin T, Li X-A, Zhu W, Uittenbogaard A, Smart EJ: HDL-associated estradiol stimulates endothelial NO synthase and vasodilation in an SR-BI-dependent manner. J. Clin. Invest. 2003, 111:1579–1587.
[12]. Hochrath K, Krawczyk M, Goebel R, Langhirt M, Rathkolb B, Micklich K, Rozman J, Horsch M, Beckers J, Klingenspor M, Fuchs H, Gailus-Durner V, Wolf E, Acalovschi M, Volmer DA, Hrabě de Angelis M, Lammert F: The hepatic phosphatidylcholine transporter ABCB4 as modulator of glucose homeostasis. FASEB J. 2012, 26:5081–5091.
[13]. Rusiñol AE, Jamil H, Vance JE: In vitro reconstitution of assembly of apolipoprotein B48-containing lipoproteins. J. Biol. Chem. 1997, 272:801
[14]. Meyer RA, Lampe PD, Malewicz B, Baumann WJ, Johnson RG: Enhanced gap junction formation with LDL and apolipoprotein B. Exp. Cell Res. 1991, 196:72–81.
[15]. Min DS, Ahn BH, Rhie DJ, Yoon SH, Hahn SJ, Kim MS, Jo YH: Expression and regulation of phospholipase D during neuronal differentiation of PC12 cells. Neuropharmacology 2001, 41:384–391.
[16]. Kwun HJ, Lee JH, Min DS, Jang KL: Transcriptional repression of cyclin-dependent kinase inhibitor p21 gene by phospholipase D1 and D2. FEBS Lett. 2003, 544:38–44.
[17]. Jiang XC, Qin S, Qiao C, Kawano K, Lin M, Skold A, Xiao X, Tall AR: Apolipoprotein B secretion and atherosclerosis are decreased in mice with phospholipid-transfer protein deficiency. Nat. Med. 2001, 7:847–852.
[18]. Tu AY, Albers JJ: Glucose regulates the transcription of human genes relevant to HDL metabolism: responsive elements for peroxisome proliferator-activated receptor are involved in the regulation of phospholipid transfer protein. Diabetes 2001, 50:1851–1856.
[19]. Masson D, Jiang X-C, Lagrost L, Tall AR: The role of plasma lipid transfer proteins in lipoprotein metabolism and atherogenesis. J. Lipid Res. 2009, 50 Suppl:S201–206.
[20]. Matsumoto S: The activities of lung stretch and irritant receptors during cough. Neurosci. Lett. 1988, 90:125–129.
[21]. Enkhmaa, Byambaa, Erdembileg Anuurad, Wei Zhang, Tina Tran, and Lars Berglund. “Lipoprotein(a): Genotype-Phenotype Relationship and Impact on Atherogenic Risk.” Metabolic Syndrome and Related Disorders 9, no. 6 (December 2011): 411–418. doi:10.1089/met.2011.0026.

Chapter 4 Discussion

4.1 Effects of smoking on metabolite profile

In the first study, the associations of metabolite concentrations with smoking were investigated using a metabolomics approach, which delineated the reversion of metabolite variations after smoking cessation. The results were demonstrated using protein-metabolite networks. Strong associations of various metabolites with smoking were identified, and thus it was possible to confirm part of the findings of the pilot study were confirmed (Wang-Sattler *et al*, 2008). Amongst the 23 smoking-related metabolites identified in the pilot study, 11 metabolites were measured in this study. Five of these 11 metabolites were validated in men as the current smokers sample set was about five-fold larger for males than for females. Consistent patterns of smoking effects on metabolite profiles were observed in the current study. Among all the smoking-related metabolites, higher unsaturated diacyl-PCs, but lower acyl-alkyl-PCs and saturated diacyl-PCs were established in current smokers, which in general indicate an increased level of unsaturated fatty acids in current smokers. The unsaturated fatty acids are more vulnerable to lipid peroxidation and influence the risk of different diseases, such as CVDs (Kris-Etherton, 1999; Mozaffarian *et al*, 2004).

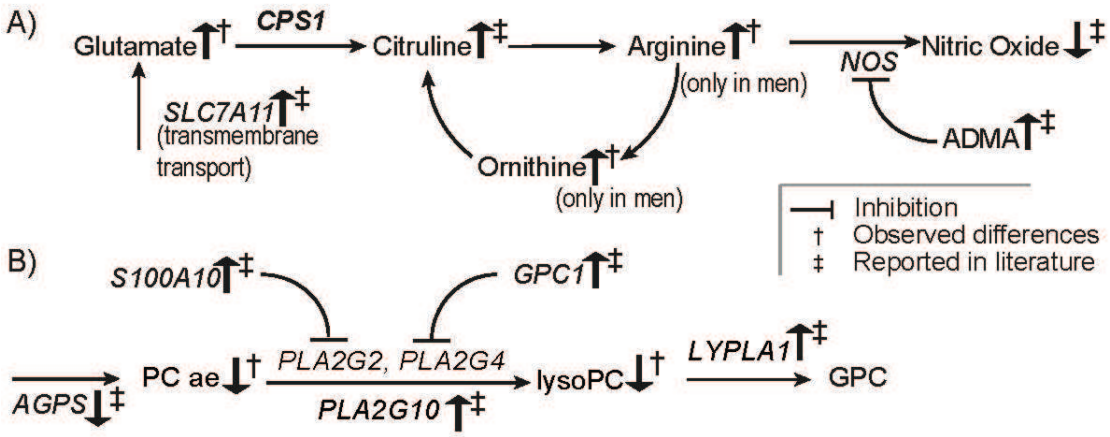

Figure 19. Pathways illustrating effects of smoking on arginine and glutamate as well as on lipid metabolism

Metabolites are in regular font, protein coding genes are in italic, gender-specific gene (CPS1) is in bold italic font. PC: phosphatidylcholine; ae: acyl-alkyl-; LYPLA1: lysophospholipase I; lysoPC: acyl-phosphatidylcholine; GPC: glycerylphosphorylcholine; ADMA: Asymmetric dimethylarginine.

From previous findings, it has been indicated that the glutamate transporter in human lung epithelial cells, encoded by the *SLC7A11* gene, is activated in current smokers (Beane *et al*, 2007; Bridges *et al*, 2001), which increases the transportation of glutamate and rises the levels of downstream metabolites arginine and ornithine subsequently (Figure 19A). The activation of cysteine-glutamate transporter (encoded by *SLC7A11*) and the increased glutamate level as a response to oxidative stress, are also of great importance to endothelial dysfunction involved at all stages of atherosclerotic plaque evolution as it leads to cardiovascular diseases (Harrison *et al*, 2003; Glass & Witztum, 2001). Ether lipid and glycerophospholipid metabolisms are associated with smoking (Wang-Sattler *et al*, 2008). As shown in Figure 19B, up regulation of *S100A10* and *GPC1* inhibits the cytosolic phospholipase A2 (cPLA2), which plays a role in the synthesis of lyso-PCs. Furthermore, the lysophospholipase I isoform, which hydrolyses lyso-PCs into glycerophosphocholine, is up regulated in current smokers (Beane *et al*, 2007), aggravating the deficiency of lyso-PCs.

The protein-metabolite interaction network shows that the reversibility of metabolite concentrations also coincided with gene expression (Figure 8). Arginine and glutamate were quickly reversed after smoking cessation, which were in line with the quick reversibility of *SLC7A11* expression. Expression of the enzyme coding genes for hydrolysis of diacyl-PCs and acyl-alkyl-PCs—for instance lysophospholipase, cPLA2, and S100 calcium binding protein A2 (S100A2)—were quickly reversible, which is consistent with our findings that the levels of diacyl-PCs and acyl-alkyl PCs quickly restored after smoking cessation (Figure 8).

In this study, the effects of smoking on metabolite profiles were found different between men and women. This result supports the hypothesis that differences in smoking effects on males and females are not solely based on smoking intensity, but are also gender-specific. Glutamate is increased in both male and female current smokers, however, the levels of arginine and ornithine only increased in male current smokers. According to the previous study of the metabolomic and genetic biomarkers on sexual dimorphisms (Mittelstrass *et al*, 2011), the *CPS1* gene, which regulates the formation of arginine, has a gender-specific manner in certain SNPs with stronger effects in women than in men. The gender-specific genetic effect might cause a lower efficiency in females with regards to the transformation of extra glutamate to citruline (Figure 19A).

In summary, a systematic targeted metabolomics approach with 140 metabolites was applied in a large population-based cohort study which examined the life-style related environmental exposure – smoking in this study. This study shows the power of metabolomics approach in investigating the molecular signature of lifestyle-related environmental exposures. It is demonstrated in this study that smoking is associated with concentration variations in amino acids, ether lipid and glycerophospholipid metabolism. The smoking-related changes in the human serum metabolite profile are reversible after stopping smoking. This indicates the remarkable benefits of smoking cessation and provides a link to CVD benefits. Furthermore, linking metabolomic knowledge to other 'omics' approaches, for example, transcriptomics, may have the potential to identify novel biomarkers as well as new risk assessment tools.

4.2 Effects of smoking on multi-level 'omics' profiles

In the second study, the complex relationships between smoking, DNA methylation, gene expression and metabolites were analyzed using mediation analyses. This approach enables integration of multi-level 'omics' data to understand the response of complex system with regard to an environmental exposure—smoking.

Significant effects of smoking on the DNA methylation and gene expression profiles in KORA F4 were discovered, which were also replicated in KORA F3. Among the 361 CpG sites found in this study, 162 of which overlapped with findings (187 CpG sites) in a previous study in KORA (Zeilinger *et al*, 2013) where a different normalization procedure was used, and thus, is considered as a proof of principle for the efficacy of the analysis. The expression levels of two genes—*LRRN3* and *CLDND1*—were significantly associated with smoking in KORA F4 and replicated in KORA F3. *LRRN3*, which encodes leucine-rich repeat neuronal protein 3, is involved in neurodevelopment. It is mapped to loci that have influence on smoking cessation behavior (Rose *et al*, 2010). Increased expression of *LRRN3* in peripheral blood in current smokers was recently reported (Wan *et al*, 2012). It has been reported (Zeilinger *et al*, 2013) and also been confirmed in this study that the *LRRN3* is relatively hypo-methylated in current smokers, while the effect is reversible after smoking cessation. This fits the model of "quickly reversible" gene expression when comparing current smokers with former smokers. The function of *LRRN3* was linked to aging and may influence the general health status of cells that are mainly involved in activation of MAPK activity and endocytosis (Ashburner *et al*, 2000). *CLDND1*, a claudin-domain containing gene, is highly expressed in lymphocytes in response to smoking (Charlesworth *et al*, 2010). I has been used as a predictive biomarker in blood for scoring of the smoking status (Beineke *et al*, 2012). It is also independently

related to CAD after adjusting for smoking status (Liu *et al*, 2007). The expression level of *CLDND1* is also reversible after smoking cessation as found in this study. However, the methlylation level of *CLDND1* was not significantly different between current smokers and never smokers, thus leaves the possibility of trans-regulation effects on the expression of *CLDND1*.

In the Mediation analyses for the associations between smoking, DNA methylation, gene expression, and metabolites, several mediation effects were found significant. Seven CpG sites were found to mediate the association between smoking and *LRRN3* gene expression. Besides cg09837977, which is within the 1M bp region of the transcriptional starting site of *LRRN3*, all other CpG sites mediated *LRRN3* expression in the trans-regulation manner. Amid all seven CpG sites, two cpg sites (cg26729380 and cg14753356) were noteworthy, as they were also significantly associated with LPC (18:2) and PC ae C34:3 levels. However, no literature support was found for the direct association between LRRN3 and the two metabolites. Further studies to proof the direct association between LRRN3 and the two metabolites are required as there was no literature to support these findings. In this study significant mediation effect of gene expression for the association between smoking and DNA methylation was also discovered, which raises the possibility of a feedback loop between expression and methylation.

Several issues warrants further investigation of these discovered mediation effects. Firstly, due to the small samples size in the replication dataset, and the drawbacks of cross-sectional studies, it is not sufficient to make a conclusive deduction from findings of this study. Secondly, while both the expression and DNA methylation profile have strong tissue specificity (Yu *et al*, 2010), whether the metabolite profile of a blood sample can reflect variations in the other two ‘omics’ levels in specific

tissues is questionable. As we have found in this study the associations between gene expression and methylation, which both reflect the same 'local' biological processes, had in general stronger associations than their associations with metabolite concentrations. For this reason, tissue-specific data are necessary to rule out the influence of other organs for the integrative study of multi-level 'omics'.

In summary, this study is a first attempt to integrate multi-level 'omics' data from large cohort studies. Using mediation analysis, the relations and interactions between methylation, gene expression and metabolite were investigated, and significant mediation effects between different 'omics' levels were discovered. These findings shows the potential of this approach to disentangle the complex relations between different 'omics' profile.

4.3 Metabolite markers of incident MI

MI is a leading cause of death worldwide (Yeh *et al*, 2010). The long latent period of the acute events provides a chance for primary prevention of the disease. A major challenge of prevention is to identify people at risk (Ajani & Ford, 2006). Despite a large number of studies screening for candidate biomarkers, the discovered risk factors of MI, such as CRP, have demonstrated significant but only modest added value to conventional risk scores, such as Framingham score for the prediction of MI (Ioannidis & Tzoulaki, 2012; C-Reactive Protein, Fibrinogen, and Cardiovascular Disease Prediction, 2012) . A Mendelian randomization study also provides negative results for the causal association between CRP and MI (C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC), 2011).

In this thesis, two population-based cohort studies were used for the investigation for the associations between metabolite concentrations and incident MI with over 2200 participants who have been followed-up for an average of nine years and have been

profiled at baseline in a metabolite panel covering lipids, amino acids and sugars metabolism. The prospective analysis allowed us to show that metabolite concentration variations occurred before the onset of MI.

Based on prospective cohort studies, five novel biomarkers for incident MI were identified by performing targeted metabolomics profiling, of which three were replicated in a prospective case-cohort study with non-fasting participants. Fasting concentration of the five biomarkers provided significant added predictive value for incident MI, but with this said, the non-fasting concentrations showed attenuated prediction. Furthermore, significant association between the metabolites and CRP were found in all cohort studies, whilst these metabolites explained for about 10% of the variations in the CRP levels.

CRP is one of the most promising biomarkers of MI. However, the Mendelian randomization study failed to prove the causality of the association between CRP and MI (C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC), 2011). In this study the attenuated association between CRP and MI was established when adding the metabolites into the association models. The results showed strong association of CRP with the identified metabolites. It raises the possibility that an inflammatory process, which is casual for the association between CRP and MI, has not yet been revealed. The metabolites may confound the association between MI and CRP, thus, can potentially server as better biomarkers for early inflammatory process in thrombosis, revealing the intrinsic biochemical process at the early stage of MI.

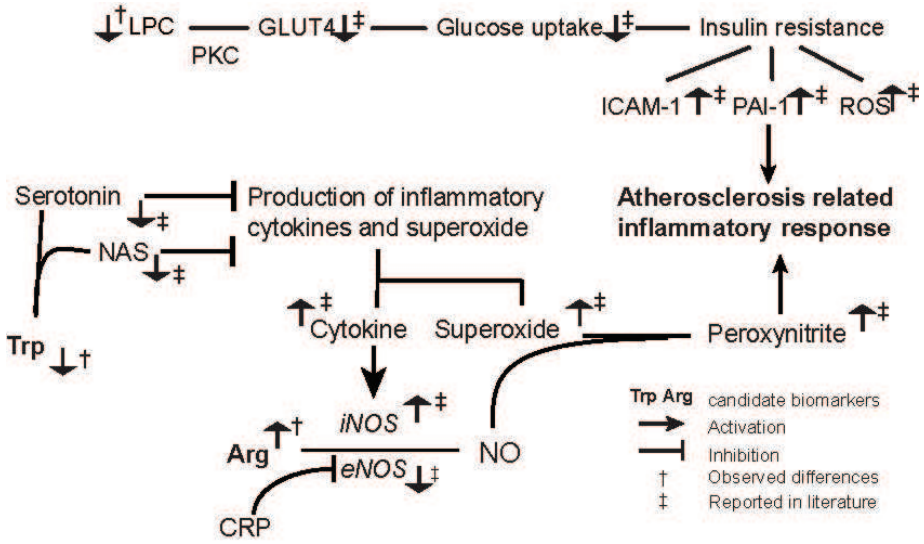

Figure 20. Potential pathways that LPC, tryptophan and arginine involved in the inflammatory response of CVDs.

Abbreviations: NAS, N-Acetylserotonin; NO, nitric oxide; iNOS, iducible nitric oxide synthase; CRP, C-reactive protein; PAI-1, Plasminogen activator inhibitor-1; ICAM-1, Intercellular Adhesion Molecule 1; ROS, Reactive oxygen species; PKC, protein kinase C.

Lipoprotein-associated phospholipase A2 and choline are providing new insight to the inflammation and disorder of lipid metabolism in MI (Wang *et al*, 2011b; Gonçalves *et al*, 2012; Shah *et al*, 2010). Our findings of the associations between incident MI and a group of lyso-PCs (LPC 16:0, LPC, 17:0, LPC 18:0, LPC 18:1, LPC (18:2)) are noteworthy in this context (Kolodgie *et al*, 2006; Gonçalves *et al*, 2012). lyso-PCs levels are higher in the plaque regions of individuals with atherosclerosis (Gonçalves *et al*, 2012). However, the lyso-PCs levels in plaque are not associated with the levels in blood (Gonçalves *et al*, 2012), implicating a tissue-specific or regional-specific manner for the variations of lyso-PCs. In our study, higher lyso-PCs levels in serum exhibited a reduced risk of MI. Lyso-PCs are associated with improved insulin sensitivity and glucose clearance (Yea *et al*, 2009). Specifically, lyso-PCs stimulate the glucose transporter type 4 (GLUT4) in adipocytes enhancing glucose uptake with the dependency on protein kinase C (Yea *et al*, 2009). With this knowledge, it can be

therefore deducted that lower levels of lyso-PCs may influence the glucose uptake in adipocytes and increase the risk of MI by increasing intercellular adhesion molecule 1 (ICAM-1), plasminogen-activator inhibitor type 1 (PAI-1) and reactive oxygen species (ROS) (Figure 20) (Van Gaal *et al*, 2006). Several other recently published studies also found lower level of plasma lyso-PCs in obesity and cardiovascular diseases which further support our findings (Barber *et al*, 2012; Fernandez *et al*, 2013).

The LPC (17:0) is potentially from ruminant origin (Nestel *et al*, 2014) or are produced by gut flora activities (Hopkins, 2001). The variations of LPC (17:0) in individuals who developed MI may result from the variation in dairy consumption or abnormality of gut flora activities in these participants, which have to have an influence on cardiovascular diseases (Huth & Park, 2012; Wang *et al*, 2011b).

The higher level of arginine and lower level of tryptophan in the people who are at high risk of developing MI, also potentially indicate early atherogenic inflammation. As shown in Figure 20, the low level of tryptophan may result in insufficiency of serotonin and NAS (Wu, 2009), which raises the levels of cytokine and superoxide in blood. Cytokine triggers the activity of inducible nitric oxide synthase (iNOS) that releases large amount of nitric oxide (NO), which in turn reacts with superoxide to generate peroxynitrite – an important component in the atherosclerosis related inflammatory response (Mungrue *et al*, 2002).

The association between tryptophan and phosphatidylcholine (PC aa C32:2) were not replicated. The meta-analysis showed significant heterogeneity between the associations in discovery and replication study (Table 16), which might reflect the differences between fasting and non-fasting samples. In the KORA S4 study, the

associations of the potential biomarkers with MI were also weaker in the sensitivity analysis combining both fasting and non-fasting samples. Dietary intake influences a large panel of metabolites in both blood and urine, including lipids and essential amino acids such as tryptophan (Krug *et al*, 2012; Hodson *et al*, 2008; Ishikawa *et al*, 2014). Food intake in general increase the inter-individual variations in metabolite concentrations (Krug *et al*, 2012), which probably adds noise for certain disease related metabolite profile.

4.4 Metabolomic linkage between smoking and MI

The results from the three studies proved the potential of metabolomics in revealing the role of an environmental factor, e.g. smoking life-style, in pathogenesis and prognosis.

In the third study of this thesis, two metabolites (arginine and LPC (18:2)) were discovered, which were related with both smoking and incident MI. Concentrations of arginine and LPC (18:2), that are associated with both smoking and coronary artery diseases, quickly returned to normal levels (within 7 years) after smoking cessation. These outcomes are in line with epidemiological findings that the smoking effect on cardiovascular disease are quickly and largely reduced after smoking cessation (Danesh *et al*, 2000, 1999, 1998). Our findings provided new insight into the metabolic basis for the reduced risk of CVD after smoking cessation and provided support for the remarkable benefits people would gain by ceasing smoking.

4.4.1 Arginine

Higher level of arginine was found in both smokers and people with MI events, which may imply a pathophysiological link between MI and smoking via this metabolite. Under physiological conditions nitric oxide synthase (NOS) oxidizes L-arginine, thereby generating L-citrulline and nitric oxide (NO). The main cellular target of NO

is the soluble guanylate cyclase (*GUCY1A3*) which catalyzes the conversion of GTP to cGMP. The formed cGMP induces vasodilation and inhibits adhesion of platelets and neutrophiles to the endothelium. Three isoforms of NOS are involved in this biological pathway, the two Ca^{2+}-dependent isoforms NOS1 (nNOS) and NOS3 (eNOS) and one cytokine-inducible isoform NOS2 (iNOS) (Knowles & Moncada, 1994).

Two scenarios provide explanations how an increased arginine level is associated with a higher risk for MI. Firstly, high arginine levels might result in an excessive generation of the arginine degradation product NO. NO interacts with superoxide anions (O.-) in the cells resulting in the formation of the cytotoxic peroxynitrite which induces tissue injury via lipid peroxidation (Wever *et al*, 1998) and inflammatory responses (Reiter *et al*, 2000). Pro-inflammatory cytokines are known to induce the expression of iNOS resulting in an even higher accumulation of NO and consequently peroxynitrite (Schulz *et al*, 1995). High arginine levels may also be a result from decreased NOS activity or impaired arginine uptake from the plasma into cells as it has been already been reported for ulcerative colitis (Hong *et al*, 2010). eNOS deficiency is a pivotal event in atherogenesis (Hong *et al*, 2010). Reduced NO-induced cGMP formation contributes to vascular contractile dysfunction, increased adhesion of immune cells to the endothelial membrane (Erdmann *et al*, 2013) and accelerated inflammatory response (Kalz *et al*, 2014).

4.4.2 LPC (18:2)

Decreased level of LPC (18:2) were found in current smokers (Wang-Sattler *et al*, 2008; Xu *et al*, 2013) and also in patients with diabetes (Wang-Sattler *et al*, 2012), which are both risk factors of MI. In smokers, the decreased level of LPC (18:2) reflects inhibition of upstream synthesis and activation of downstream hydrolysis.

Interestingly, one recent study showed that a disorder of phosphatidylcholine metabolism would promote cardiovascular disease (Wang *et al*, 2011b), thus a link might be established between smoking-related phosphatidylcholine variation and cardiovascular events. For example, the phosphatidylcholine hydroperoxide will promote angiogenesis in endothelial cells that are associated with atherosclerotic development (Nakagawa *et al*, 2011). Lyso-PCs show pro- and antiatherogenic effects (Schmitz & Ruebsaamen, 2010). Lyso-PCs enhance the transcriptional activation of endothelial nitric-oxide synthase (eNOS, NOS3) gene (Cieslik *et al*, 1998). Lyso-PC significantly increased the expression of extracellular-superoxide dismutase SOD3 mRNA and protein in human monocytic U937 cells. Since SOD3 is a scavenger of superoxide, the up-regulation of SOD3 could lead to protection of the biological activity of NO and blocking peroxynitrite toxicity (Yamamoto *et al*, 2002).

4.5 Conclusions and future prospects

In this thesis, three projects were presented exploring smoking, MI and their relation using 'omics' data from large population-based cohort studies. The results link the life-style related environmental exposure, thus smoking, and the link between smoking and disease outcome: MI on a molecular level. In the first project, significant variations of metabolite concentration in smokers were found, which were reversible after smoking cessation. In the second project, using a systems biology approach that integrated transcriptomic, epigenomic and metabolite profiling data, two CpG sites and *LRRN3* were discovered as mediators for the change in the concentration of LPC (18:2) and PC ae C34:3 under smoking exposure. In the third project, three novel metabolites associated with incident MI were identified in prospective cohort studies. Two metabolites, arginine and LPC (18:2) were associated with both smoking and MI, underlined metabolic basis for the association between the two phenotypes. These

metabolites were significantly related to early inflammatory processes as shown in network analysis and their significant relations with inflammatory markers, such as CRP.

Several issues warrant further investigation. In all three projects, the analyses were based on a limited range and number of metabolites, which cannot fully represent the whole metabolome. Although lipid metabolism and amino acids pathways were mainly considered in the study of cardiovascular diseases, a more comprehensive metabolite profile will help to discover novel biomarkers in new pathways. Thus, an improved metabolomics technic measuring more metabolites is urgently needed to get a comprehensive understanding of the effects of smoking and MI on the human metabolism. It would be interesting for future studies to also include data on other environmental factors such as diet and life style which are known to have effects on the human metabolism (Gibney *et al*, 2005; Pohjanen *et al*, 2007).

A larger prospective study is warranted for replication of our results. For the second project, the study does not have sufficient power for mediation analysis in a genome wide level. The samples size in the cohort for replication was too small to draw any concrete conclusions from our discovery. A similar issue also exists in the third project, in which both the discovery and replication cohort studies have relative small number of incident MI cases. Also, subgroup analysis will not be sufficiently powered under this context. Metabolomic study has shown significant differences of metabolite profile between men and women, which implies that there are gender-specific genes at play which leads to crucial variations in metabolite profiles (Mittelstrass *et al*, 2011). A large prospective study will help to empower the analyses of subgroups or interactions between different factors, which may help to selectively use predictive biomarkers in subpopulation.

References

Aberg A, Bergstrand R, Johansson S, Ulvenstam G, Vedin A, Wedel H, Wilhelmsson C & Wilhelmsen L (1983) Cessation of smoking after myocardial infarction. Effects on mortality after 10 years. *Br Heart J* **49:** 416–422

Ajani UA & Ford ES (2006) Has the risk for coronary heart disease changed among U.S. adults? *J. Am. Coll. Cardiol.* **48:** 1177–1182

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19:** 716–723

Altmaier E, Kastenmüller G, Römisch-Margl W, Thorand B, Weinberger KM, Illig T, Adamski J, Döring A & Suhre K (2011) Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur. J. Epidemiol.* **26:** 145–156

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM & Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25:** 25–29

Baker M (2011) Metabolomics: from small molecules to big ideas. *Nat Meth* **8:** 117–121

Bakhru A & Erlinger TP (2005) Smoking cessation and cardiovascular disease risk factors: results from the Third National Health and Nutrition Examination Survey. *PLoS Med.* **2:** e160

Barabási A-L, Gulbahce N & Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nature Reviews Genetics* **12:** 56–68

Barber MN, Risis S, Yang C, Meikle PJ, Staples M, Febbraio MA & Bruce CR (2012) Plasma Lysophosphatidylcholine Levels Are Reduced in Obesity and Type 2 Diabetes. *PLoS ONE* **7:** e41456

Barlow WE, Ichikawa L, Rosner D & Izumi S (1999) Analysis of case-cohort designs. *J Clin Epidemiol* **52:** 1165–1172

Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME & Spira A (2007) Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol.* **8:** R201

Becker RA & Chambers JM (1984) S: an interactive environment for data analysis and graphics Belmont, Calif.: Wadsworth Advanced Book Program

Beineke P, Fitch K, Tao H, Elashoff MR, Rosenberg S, Kraus WE, Wingrove JA & $author.lastName $author firstName (2012) A whole blood gene expression-based signature for smoking status. *BMC Medical Genomics* **5:** 58

Bellavia A, Urch B, Speck M, Brook RD, Scott JA, Albetti B, Behbod B, North M, Valeri L, Bertazzi PA, Silverman F, Gold D & Baccarelli AA (2013) DNA Hypomethylation, Ambient Particulate Matter, and Increased Blood Pressure: Findings From Controlled Human Exposure Experiments. *J Am Heart Assoc* **2:** e000212

Bolstad BM, Irizarry RA, Astrand M & Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19:** 185–193

Bøttcher M & Falk E (1999) Pathology of the coronary arteries in smokers and non-smokers. *J Cardiovasc Risk* **6:** 299–302

Bridges CC, Kekuda R, Wang H, Prasad PD, Mehta P, Huang W, Smith SB & Ganapathy V (2001) Structure, function, and regulation of human cystine/glutamate transporter in retinal pigment epithelial cells. *Invest. Ophthalmol. Vis. Sci.* **42:** 47–54

C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC) (2011) Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ* **342:** d548–d548

Carter SL, Brechbühler CM, Griffin M & Bond AT (2004) Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20:** 2242–2250

Cavill R, Kamburov A, Ellis JK, Athersuch TJ, Blagrove MSC, Herwig R, Ebbels TMD & Keun HC (2011) Consensus-Phenotype Integration of Transcriptomic and Metabolomic Data Implies a Role for Metabolism in the Chemosensitivity of Tumour Cells. *PLoS Comput Biol* **7:** Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3068923/ [Accessed April 12, 2014]

Chan SY & Loscalzo J (2012) The Emerging Paradigm of Network Medicine in the Study of Human Disease. *Circulation Research* **111:** 359–374

Charlesworth JC, Curran JE, Johnson MP, Göring HH, Dyer TD, Diego VP, Kent JW, Mahaney MC, Almasy L, MacCluer JW, Moses EK & Blangero J (2010) Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. *BMC Medical Genomics* **3:** 29

Chen Y, Chen C, Ke X, Xiong L, Shi Y, Li J, Tan X & Ye S (2014) Analysis of Circulating Cholesterol Levels as a Mediator of an Association between ABO Blood Group and Coronary Heart Disease. *Circ Cardiovasc Genet***:** CIRCGENETICS.113.000299

Chuang H-Y, Hofree M & Ideker T (2010) A Decade of Systems Biology. *Annual Review of Cell and Developmental Biology* **26:** 721–744

Cieslik K, Zembowicz A, Tang JL & Wu KK (1998) Transcriptional regulation of endothelial nitric-oxide synthase by lysophosphatidylcholine. *J. Biol. Chem.* **273:** 14885–14890

Clinger W & Rees J (1991) The revised4 report on the algorithmic language Scheme. *ACM Lisp Pointers* **4:** 1–55

Consortium TT and G (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42:** 441–447

Cook DG, Shaper AG, Pocock SJ & Kussick SJ (1986) Giving up smoking and the risk of heart attacks. A report from The British Regional Heart Study. *Lancet* **2:** 1376–1380

C-Reactive Protein, Fibrinogen, and Cardiovascular Disease Prediction (2012) *New England Journal of Medicine* **367:** 1310–1320

Csardi G & Nepusz T (2006) The igraph software package for complex network research. *InterJournal* **Complex Systems:** 1695

D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM & Kannel WB (2008) General Cardiovascular Risk Profile for Use in Primary Care The Framingham Heart Study. *Circulation* **117:** 743–753

Danesh J, Collins R, Appleby P & Peto R (1998) Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA* **279:** 1477–1482

Danesh J, Muir J, Wong YK, Ward M, Gallimore JR & Pepys MB (1999) Risk factors for coronary heart disease and acute-phase proteins. A population-based study. *Eur. Heart J.* **20:** 954–959

Danesh J, Whincup P, Walker M, Lennon L, Thomson A, Appleby P, Gallimore JR & Pepys MB (2000) Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. *BMJ* **321:** 199–204

Davey Smith G (2011) Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. *Genes Nutr* **6:** 27–43

Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, König IR, Cazier J-B, Johansson Å, Hall AS, Lee J-Y, Willer CJ, Chambers JC, Esko T, Folkersen L, et al (2012) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics* **45:** 25–33

Dijkstra EW (1959) A note on two problems in connexion with graphs. *Numerische Mathematik* **1:** 269–271

Du P, Kibbe WA & Lin SM (2008) lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24:** 1547–1548

Ebbert JO, Yang P, Vachon CM, Vierkant RA, Cerhan JR, Folsom AR & Sellers TA (2003) Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women. *J. Clin. Oncol.* **21:** 921–926

Ellis JK, Athersuch TJ, Thomas LD, Teichert F, Pérez-Trujillo M, Svendsen C, Spurgeon DJ, Singh R, Järup L, Bundy JG & Keun HC (2012) Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Med* **10:** 61

Erdmann J, Stark K, Esslinger UB, Rumpf PM, Koesling D, de Wit C, Kaiser FJ, Braunholz D, Medack A, Fischer M, Zimmermann ME, Tennstedt S, Graf E, Eck S, Aherrahrou Z, Nahrstaedt J, Willenborg C, Bruse P, Brænne I, Nöthen MM, et al (2013) Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature* **504:** 432–436

Fernandez C, Sandin M, Sampaio JL, Almgren P, Narkiewicz K, Hoffmann M, Hedner T, Wahlstrand B, Simons K, Shevchenko A, James P & Melander O (2013) Plasma Lipid Composition and Risk of Developing Cardiovascular Disease. *PLoS ONE* **8:** e71846

Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost H-G, Fritsche A, Häring H-U, Angelis MH de, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H & Pischon T (2012) Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. *Diabetes***:** DB_120495

Freathy RM, Kazeem GR, Morris RW, Johnson PCD, Paternoster L, Ebrahim S, Hattersley AT, Hill A, Hingorani AD, Holst C, Jefferis BJ, Kring SII, Mooser V, Padmanabhan S, Preisig M, Ring SM, Sattar N, Upton MN, Vollenweider P, Waeber G, et al (2011) Genetic variation at CHRNA5-CHRNA3-CHRNB4 interacts with smoking status to influence body mass index. *Int J Epidemiol* **40:** 1617–1628

Functional genomics: lessons from yeast (2002) *Philosophical Transactions of the Royal Society B: Biological Sciences* **357:** 17–23

Van Gaal LF, Mertens IL & De Block CE (2006) Mechanisms linking obesity with cardiovascular disease. *Nature* **444:** 875–880

Gibney MJ, Walsh M, Brennan L, Roche HM, German B & van Ommen B (2005) Metabolomics in human nutrition: opportunities and challenges. *Am. J. Clin. Nutr.* **82:** 497–503

Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, Meitinger T, Mewes H-W, Wichmann H-E, Weinberger KM, Adamski J, Illig T & Suhre K (2008) Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite Profiles in Human Serum. *PLoS Genet* **4:** e1000282

Glass CK & Witztum JL (2001) Atherosclerosis. the road ahead. *Cell* **104:** 503–516

Gonçalves I, Edsfeldt A, Ko NY, Grufman H, Berg K, Björkbacka H, Nitulescu M, Persson A, Nilsson M, Prehn C, Adamski J & Nilsson J (2012) Evidence Supporting a Key Role of Lp-PLA2-Generated Lysophosphatidylcholine in Human Atherosclerotic Plaque Inflammation. *Arterioscler Thromb Vasc Biol* **32:** 1505–1512

Guo M, House MG, Hooker C, Han Y, Heath E, Gabrielson E, Yang SC, Baylin SB, Herman JG & Brock MV (2004) Promoter hypermethylation of resected bronchial margins: a field defect of changes? *Clin. Cancer Res.* **10:** 5131–5136

Halpern MT, Gillespie BW & Warner KE (1993) Patterns of absolute risk of lung cancer mortality in former smokers. *J. Natl. Cancer Inst.* **85:** 457–464

Harrison D, Griendling KK, Landmesser U, Hornig B & Drexler H (2003) Role of oxidative stress in atherosclerosis. *The American Journal of Cardiology* **91:** 7–11

Hodson L, Skeaff CM & Fielding BA (2008) Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Prog. Lipid Res.* **47:** 348–380

Holle R, Happich M, Löwel H & Wichmann HE (2005) KORA--a research platform for population based health research. *Gesundheitswesen* **67 Suppl 1:** S19–25

Hong M-G, Myers AJ, Magnusson PKE & Prince JA (2008) Transcriptome-Wide Assessment of Human Brain and Lymphocyte Senescence. *PLoS ONE* **3:** e3024

Hong S-KS, Maltz BE, Coburn LA, Slaughter JC, Chaturvedi R, Schwartz DA & Wilson KT (2010) Increased serum levels of L-arginine in ulcerative colitis and correlation with disease severity. *Inflamm. Bowel Dis.* **16:** 105–111

Hopkins MJ (2001) Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut* **48:** 198–205

Howard G, Burke GL, Szklo M, Tell GS, Eckfeldt J, Evans G & Heiss G (1994) Active and passive smoking are associated with increased carotid wall thickness. The Atherosclerosis Risk in Communities Study. *Arch. Intern. Med.* **154:** 1277–1282

Huelsenbeck JP & Crandall KA (1997) Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood. *Annual Review of Ecology and Systematics* **28:** 437–466

Huth PJ & Park KM (2012) Influence of Dairy Product and Milk Fat Consumption on Cardiovascular Disease Risk: A Review of the Evidence. *Adv Nutr* **3:** 266–285

Ideker T, Galitski T & Hood L (2001) A NEW APPROACH TO DECODING LIFE: Systems Biology. *Annual Review of Genomics and Human Genetics* **2:** 343–372

Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmüller G, Kato BS, Mewes H-W, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann H-E, Spector TD, Adamski J & Suhre K (2009) A genome-wide perspective of genetic variation in human metabolism. *Nature Genetics* **42:** 137–141

Ioannidis JPA & Tzoulaki I (2012) Minimal and Null Predictive Effects for the Most Popular Blood Biomarkers of Cardiovascular Disease. *Circulation Research* **110:** 658–662

Ishikawa M, Maekawa K, Saito K, Senoo Y, Urata M, Murayama M, Tajima Y, Kumagai Y & Saito Y (2014) Plasma and Serum Lipidomics of Healthy White Adults Shows Characteristic Profiles by Subjects' Gender and Age. *PLoS ONE* **9:** e91806

Jourdan C, Petersen A-K, Gieger C, Döring A, Illig T, Wang-Sattler R, Meisinger C, Peters A, Adamski J, Prehn C, Suhre K, Altmaier E, Kastenmüller G, Römisch-Margl W, Theis FJ, Krumsiek J, Wichmann H-E & Linseisen J (2012) Body Fat Free Mass Is Associated with the Serum Metabolite Profile in a Population-Based Study. *PLoS ONE* **7:** e40009

Kalinina OV, Wichmann O, Apic G & Russell RB (2011) Combinations of Protein-Chemical Complex Structures Reveal New Targets for Established Drugs. *PLoS Comput Biol* **7:** e1002043

Kalz J, ten Cate H & Spronk HMH (2014) Thrombin generation and atherosclerosis. *J. Thromb. Thrombolysis* **37:** 45–55

Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28:** 27–30

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M & Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42:** D199–205

Kastenmüller G, Römisch-Margl W, Wägele B, Altmaier E & Suhre K (2011) metaP-Server: A Web-Based Metabolomics Data Analysis Tool. *Journal of Biomedicine and Biotechnology* **2011:** 1–8

Kitano H (2002) Systems Biology: A Brief Overview. *Science* **295:** 1662–1664

Knolhoff AM, Nautiyal KM, Nemes P, Kalachikov S, Morozova I, Silver R & Sweedler JV (2013) Combining small-volume metabolomic and transcriptomic approaches for assessing brain chemistry. *Anal. Chem.* **85:** 3136–3143

Knowles RG & Moncada S (1994) Nitric oxide synthases in mammals. *Biochem. J.* **298 (Pt 2):** 249–258

Kolodgie FD, Burke AP, Skorija KS, Ladich E, Kutys R, Makuria AT & Virmani R (2006) Lipoprotein-associated phospholipase A2 protein expression in the

natural progression of human coronary atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **26:** 2523–2529

Kris-Etherton PM (1999) AHA Science Advisory. Monounsaturated fatty acids and risk of cardiovascular disease. American Heart Association. Nutrition Committee. *Circulation* **100:** 1253–1258

Krug S, Kastenmüller G, Stückler F, Rist MJ, Skurk T, Sailer M, Raffler J, Römisch-Margl W, Adamski J, Prehn C, Frank T, Engel K-H, Hofmann T, Luy B, Zimmermann R, Moritz F, Schmitt-Kopplin P, Krumsiek J, Kremer W, Huber F, et al (2012) The dynamic range of the human metabolome revealed by challenges. *FASEB J* **26:** 2607–2619

Kuhn M, Szklarczyk D, Franceschini A, von Mering C, Jensen LJ & Bork P (2012) STITCH 3: zooming in on protein-chemical interactions. *Nucleic Acids Research* **40:** D876–D880

Langfelder P & Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9:** 559

Liang K-Y & Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* **73:** 13–22

Libby P (2002) Inflammation in atherosclerosis. *Nature* **420:** 868–874

Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekström TJ & Feinberg AP (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotech* **31:** 142–147

Liu Y, Sun W, Zhang K, Zheng H, Ma Y, Lin D, Zhang X, Feng L, Lei W, Zhang Z, Guo S, Han N, Tong W, Feng X, Gao Y & Cheng S (2007) Identification of genes differentially expressed in human primary lung squamous cell carcinoma. *Lung Cancer* **56:** 307–317

Long JZ, Cisar JS, Milliken D, Niessen S, Wang C, Trauger SA, Siuzdak G & Cravatt BF (2011) Metabolomics annotates ABHD3 as a physiologic regulator of medium-chain phospholipids. *Nat Chem Biol* **7:** 763–765

Lusis AJ & Weiss JN (2010) Cardiovascular Networks Systems-Based Approaches to Cardiovascular Disease. *Circulation* **121:** 157–170

De Maat MP, Pietersma A, Kofflard M, Sluiter W & Kluft C (1996) Association of plasma fibrinogen levels with coronary artery disease, smoking and inflammatory markers. *Atherosclerosis* **121:** 185–191

Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engström G, Östling G, Clish C, Wang TJ, Gerszten RE & Melander O (2013) A diabetes-predictive amino acid score and future cardiovascular disease. *Eur Heart J* **34:** 1982–1989

Meslamani J & Rognan D (2011) Enhancing the accuracy of chemogenomic models with a three-dimensional binding site kernel. *J Chem Inf Model* **51:** 1593–1603

Mittelstrass K, Ried JS, Yu Z, Krumsiek J, Gieger C, Prehn C, Roemisch-Margl W, Polonikov A, Peters A, Theis FJ, Meitinger T, Kronenberg F, Weidinger S, Wichmann HE, Suhre K, Wang-Sattler R, Adamski J & Illig T (2011) Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **7:** e1002215

Moreau Y & Tranchevent L-C (2012) Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nat Rev Genet* **13:** 523–536

Mozaffarian D, Rimm EB & Herrington DM (2004) Dietary fats, carbohydrate, and progression of coronary atherosclerosis in postmenopausal women. *Am. J. Clin. Nutr.* **80:** 1175–1184

Munafò MR & Araya R (2010) Cigarette smoking and depression: a question of causation. *BJP* **196:** 425–426

Mungrue IN, Husain M & Stewart DJ (2002) The role of NOS in heart failure: lessons from murine genetic models. *Heart Fail Rev* **7:** 407–422

Nabel EG & Braunwald E (2012) A Tale of Coronary Artery Disease and Myocardial Infarction. *New England Journal of Medicine* **366:** 54–63

Nakagawa K, Shibata A, Saito T, Sookwong P, Kato S, Tsuduki T, Matsubara K & Miyazawa T (2011) Phosphatidylcholine hydroperoxide promotes VEGF-induced angiogenesis in endothelial cells and rat aorta ring cultures. *Biochim. Biophys. Acta* **1810:** 1205–1211

Nestel PJ, Straznicky N, Mellett NA, Wong G, Souza DPD, Tull DL, Barlow CK, Grima MT & Meikle PJ (2014) Specific plasma lipid classes and phospholipid fatty acids indicative of dairy food consumption associate with insulin sensitivity. *Am J Clin Nutr***:** ajcn.071712

Patti GJ, Yanes O & Siuzdak G (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13:** 263–269

Pearce N (2012) Classification of epidemiological study designs. *International Journal of Epidemiology* **41:** 393–397

Pedersen ER, Svingen GFT, Schartum-Hansen H, Ueland PM, Ebbing M, Nordrehaug JE, Igland J, Seifert R, Nilsen RM & Nygård O (2013) Urinary excretion of kynurenine and tryptophan, cardiovascular events, and mortality after elective coronary angiography. *Eur Heart J* **34:** 2689–2696

Petersen A-K, Zeilinger S, Kastenmüller G, Römisch-Margl W, Brugger M, Peters A, Meisinger C, Strauch K, Hengstenberg C, Pagel P, Huber F, Mohney RP, Grallert H, Illig T, Adamski J, Waldenberger M, Gieger C & Suhre K (2014) Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum. Mol. Genet.* **23:** 534–545

Pohjanen E, Thysell E, Jonsson P, Eklund C, Silfver A, Carlsson I-B, Lundgren K, Moritz T, Svensson MB & Antti H (2007) A multivariate screening strategy for investigating metabolic effects of strenuous physical exercise in human serum. *J. Proteome Res.* **6:** 2113–2120

Powell CA, Klares S, O'Connor G & Brody JS (1999) Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. *Clin. Cancer Res.* **5:** 2025–2034

Pradhan AD, Manson JE, Rossouw JE, Siscovick DS, Mouton CP, Rifai N, Wallace RB, Jackson RD, Pettinger MB & Ridker PM (2002) Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative observational study. *JAMA* **288:** 980–987

Prentice RL (1986) A Case-Cohort Design for Epidemiologic Cohort Studies and Disease Prevention Trials. *Biometrika* **73:** 1

R Core Team (2013) R: A Language and Environment for Statistical Computing Vienna, Austria: R Foundation for Statistical Computing Available at: http://www.R-project.org

Reiter CD, Teng RJ & Beckman JS (2000) Superoxide reacts with nitric oxide to nitrate tyrosine at physiological pH via peroxynitrite. *J. Biol. Chem.* **275:** 32460–32466

Rhee EP & Gerszten RE (2012) Metabolomics and Cardiovascular Biomarker Discovery. *Clinical Chemistry* **58:** 139–147

Rigotti NA & Pasternak RC (1996) Cigarette smoking and coronary heart disease: risks and management. *Cardiol Clin* **14:** 51–68

Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K & Adamski J (2012) Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* **8:** 133–142

Rose JE, Behm FM, Drgon T, Johnson C & Uhl GR (2010) Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score. *Mol. Med.* **16:** 247–253

Rothman KJ (2012) Epidemiology: An Introduction Oxford University Press

Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, GuhaThakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, et al (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* **37:** 710–717

Schmerler D, Neugebauer S, Ludewig K, Bremer-Streck S, Brunkhorst FM & Kiehntopf M (2012) Targeted metabolomics for discrimination of systemic inflammatory disorders in critically ill patients. *J. Lipid Res.* **53:** 1369–1375

Schmitz G & Ruebsaamen K (2010) Metabolism and atherogenic disease association of lysophosphatidylcholine. *Atherosclerosis* **208:** 10–18

Schnabel RB, Baccarelli A, Lin H, Ellinor PT & Benjamin EJ (2012) Next Steps in Cardiovascular Disease Genomic Research—Sequencing, Epigenetics, and Transcriptomics. *Clinical Chemistry* **58:** 113–126

Schulz R, Panas DL, Catena R, Moncada S, Olley PM & Lopaschuk GD (1995) The role of nitric oxide in cardiac depression induced by interleukin-1 beta and tumour necrosis factor-alpha. *Br. J. Pharmacol.* **114:** 27–34

Sebastiani P, Nolan VG, Baldwin CT, Abad-Grau MM, Wang L, Adewoye AH, McMahon LC, Farrer LA, Taylor JG 6th, Kato GJ, Gladwin MT & Steinberg MH (2007) A network model to predict the risk of death in sickle cell disease. *Blood* **110:** 2727–2735

Shah SH, Bain JR, Muehlbauer MJ, Stevens RD, Crosslin DR, Haynes C, Dungan J, Newby LK, Hauser ER, Ginsburg GS, Newgard CB & Kraus WE (2010) Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. *Circ Cardiovasc Genet* **3:** 207–214

Shah SH, Kraus WE & Newgard CB (2012) Metabolomic Profiling for the Identification of Novel Biomarkers and Mechanisms Related to Common Cardiovascular Diseases Form and Function. *Circulation* **126:** 1110–1120

Smith GD & Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32:** 1–22

Smoot ME, Ono K, Ruscheinski J, Wang P-L & Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27:** 431–432

Sobel M (1982) Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology* **13:** 290–312

Suhre K, Shin S-Y, Petersen A-K, Mohney RP, Meredith D, Wägele B, Altmaier E, CARDIoGRAM, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes H-W, et al (2011) Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477:** 54–60

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ & von Mering C (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39:** D561–568

Szmitko PE, Wang C-H, Weisel RD, Jeffries GA, Anderson TJ & Verma S (2003) Biomarkers of vascular disease linking inflammation to endothelial activation: Part II. *Circulation* **108:** 2041–2048

Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D & Beck S (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29:** 189–196

Thomas PD, Mi H, Swan GE, Lerman C, Benowitz N, Tyndale RF, Bergen AW, Conti DV & Pharmacogenetics of Nicotine Addiction and Treatment Consortium (2009) A systems biology network model for genetic association studies of nicotine addiction and treatment. *Pharmacogenet. Genomics* **19:** 538–551

Tyson JJ, Chen K & Novak B (2001) Network dynamics and cell physiology. *Nat Rev Mol Cell Biol* **2:** 908–916

Vidal M, Cusick ME & Barabási A-L (2011) Interactome Networks and Human Disease. *Cell* **144:** 986–998

Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, Hindy G, Hólm H, Ding EL, Johnson T, Schunkert H, Samani NJ, Clarke R, Hopewell JC, Thompson JF, Li M, Thorleifsson G, Newton-Cheh C, Musunuru K, Pirruccello JP, et al (2012) Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380:** 572–580

Vulimiri SV, Misra M, Hamm JT, Mitchell M & Berger A (2009) Effects of mainstream cigarette smoke on the global metabolome of human lung epithelial cells. *Chem. Res. Toxicol.* **22:** 492–503

Wahl S, Holzapfel C, Yu Z, Breier M, Kondofersky I, Fuchs C, Singmann P, Prehn C, Adamski J, Grallert H, Illig T, Wang-Sattler R & Reinehr T (2013) Metabolomics reveals determinants of weight loss during lifestyle intervention in obese children. *Metabolomics* **9:** 1157–1167

Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson W, Lomas DA & DeMeo DL (2012) Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum. Mol. Genet.* **21:** 3073–3082

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB & Gerszten RE (2011a) Metabolite profiles and the risk of developing diabetes. *Nature Medicine* **17:** 448–453

Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung Y-M, Wu Y, Schauer P, Smith JD, Allayee H, Tang WHW, DiDonato JA, Lusis AJ & Hazen SL (2011b) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472:** 57–63

Wang-Sattler R, Yu Y, Mittelstrass K, Lattka E, Altmaier E, Gieger C, Ladwig KH, Dahmen N, Weinberger KM, Hao P, Liu L, Li Y, Wichmann H-E, Adamski J, Suhre K & Illig T (2008) Metabolic profiling reveals distinct variations linked

to nicotine consumption in humans--first results from the KORA study. *PLoS ONE* **3:** e3863

Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, Xu T, Bader E, Huth C, Mittelstrass K, Döring A, Meisinger C, Gieger C, Prehn C, Roemisch-Margl W, et al (2012) Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol.* **8:** 615

Wardle J, Carnell S, Haworth CMA, Farooqi IS, O'Rahilly S & Plomin R (2008) Obesity Associated Genetic Variation in *FTO* Is Associated with Diminished Satiety. *The Journal of Clinical Endocrinology & Metabolism* **93:** 3640–3643

Weinberger K & Graber A (2005) Using comprehensive metabolomics to identify novel biomarkers. *Screening Trends in Drug Discovery* **6:** 42–45

Weinberger KM (2008) Einsatz von Metabolomics zur Diagnose von Stoffwechselkrankheiten. *Therapeutische Umschau* **65:** 487–491

Wever RM, Lüscher TF, Cosentino F & Rabelink TJ (1998) Atherosclerosis and the two faces of endothelial nitric oxide synthase. *Circulation* **97:** 108–112

Wichmann H-E, Gieger C & Illig T (2005) KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1:** S26–30

Wilmes A, Limonciel A, Aschauer L, Moenks K, Bielow C, Leonard MO, Hamon J, Carpi D, Ruzek S, Handler A, Schmal O, Herrgen K, Bellwon P, Burek C, Truisi GL, Hewitt P, Di Consiglio E, Testai E, Blaauboer BJ, Guillou C, et al (2013) Application of integrated transcriptomic, proteomic and metabolomic profiling for the delineation of mechanisms of drug induced cell stress. *J Proteomics* **79:** 180–194

Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, et al (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37:** D603–610

Wistuba II, Lam S, Behrens C, Virmani AK, Fong KM, LeRiche J, Samet JM, Srivastava S, Minna JD & Gazdar AF (1997) Molecular damage in the bronchial epithelium of current and former smokers. *J. Natl. Cancer Inst.* **89:** 1366–1373

Wistuba II, Mao L & Gazdar AF (2002) Smoking molecular damage in bronchial epithelium. *Oncogene* **21:** 7298–7306

Woodward M (1999) Epidemiology: Study Design and Data Analysis CRC Press

Wu G (2009) Amino acids: metabolism, functions, and nutrition. *Amino Acids* **37:** 1–17

Xia J, Mandal R, Sinelnikov IV, Broadhurst D & Wishart DS (2012) MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* **40:** W127–133

Xia J, Psychogios N, Young N & Wishart DS (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **37:** W652–660

Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, Perstorfer K, Jaremek M, Roemisch-Margl W, Rathmann W, Li Y, Wichmann H-E, Wallaschofski H, Ladwig KH, Theis F, Suhre K, Adamski J, Illig T, Peters A & Wang-Sattler R (2013) Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med* **11:** 60

Yamamoto M, Hara H & Adachi T (2002) The expression of extracellular-superoxide dismutase is increased by lysophosphatidylcholine in human monocytic U937 cells. *Atherosclerosis* **163:** 223–228

Yanbaeva DG, Dentener MA, Creutzberg EC, Wesseling G & Wouters EFM (2007) Systemic effects of smoking. *Chest* **131:** 1557–1566

Yea K, Kim J, Yoon JH, Kwon T, Kim JH, Lee BD, Lee H-J, Lee SJ, Kim J-I, Lee TG, Baek M-C, Park HS, Park KS, Ohba M, Suh P-G & Ryu SH (2009) Lysophosphatidylcholine Activates Adipocyte Glucose Uptake and Lowers Blood Glucose Levels in Murine Models of Diabetes. *J. Biol. Chem.* **284:** 33833–33840

Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV & Go AS (2010) Population Trends in the Incidence and Outcomes of Acute Myocardial Infarction. *New England Journal of Medicine* **362:** 2155–2165

Yu Y, Xu T, Yu Y, Hao P & Li X (2010) Association of tissue lineage and gene expression: conservatively and differentially expressed genes define common and special functions of tissues. *BMC Bioinformatics* **11:** S1

Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, Römisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, Standl M, Thiering E, Mittelstraß K, Wichmann HE, Peters A, Suhre K, Li Y, Adamski J, Spector TD, et al (2012) Human Serum Metabolic Profiles Are Age Dependent. *Aging cell* Available at: http://www.ncbi.nlm.nih.gov/pubmed/22834969 [Accessed August 8, 2012]

Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M & Illig T (2013) Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS ONE* **8:** e63812

Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, Tu Z, Brem RB, Bumgarner RE & Schadt EE (2012) Stitching together Multiple Data Dimensions Reveals Interacting Metabolomic and Transcriptomic Networks That Modulate Cell Regulation. *PLoS Biology* **10:** e1001301

Appendix

Table S1 Metabolite panels of KORA S4 and MONICA S2 case-cohort

Table S2 CpG sites different between S and never smokers in KORA F4 and replicated in KORA F3

Table S3 Mediation analysis of smoking associated methylation with gene expression as mediator in F4 and F3

Table S2 Metabolite panels of KORA S4 and MONICA S2 case-cohort

		Biocrates Absolute IDQ P180										Biocrates Absolute IDQ P150				
		KORA S4					MONICA/KORA S2					KORA F4				
Abbreviation	**Biochemical name**	**CV (%)**	**Above LOD (%)**		**Concentration**	**Application**	**CV (%)**	**Above LOD (%)**		**Concentration**	**Application**	**r**	**Above LOD (%)**	**Mean (µM)**	**CV**	**Application**
C0	Carnitine	5.8	99.63		41.25 ± 8.99	Used	9.41	100		43.74±9.66	Used	0.88	100	35.699	6.70%	Used
C2	Acetylcarnitine	6.3	99.63		8.64 ± 2.91	Used	5.83	100		7.36±2.81	Used	0.94	100	8.223	9.40%	Used
C3	Propionylcarnitine	10	99.63		0.48 ± 0.15	Used	7.49	100		0.51±0.2	Used	0.86	100	0.397	8.00%	Used
C3:1	Propenoylcarnitine	32.8	3.72		0.01 ± 0	Excluded	24.63	6.2		0.02±0	Excluded	-0.11	0.1	0.01	37.50%	Excluded
C3-OH	Hydroxypropionylcarnitine	44.7	2.85		0.12 ± 0.1	Excluded	63.57	0		0.08±0.05	Excluded	0.05	0.36	0.03	76.60%	Excluded
C4	Butyrylcarnitine	9.7	99.63		0.23 ± 0.13	Used	8.11	100		0.21±0.09	Used	0.89	100	0.227	8.80%	Used
C4:1	Butenylcarnitine	22.2	46.25		0.02 ± 0.01	Excluded	17.88	9.43		0.03±0.01	Excluded	0.04	5.65	0.018	34.70%	Excluded
C4-OH (C3-DC)	Hydroxybutyrylcarnitine	21.1	18.95		0.06 ± 0.03	Excluded	18.82	50.65		0.05±0.03	Used	0.47	8.4	0.098	35.50%	Excluded
C5	Valerylcarnitine	10.8	98.7		0.17 ± 0.11	Used	8.51	99.87		0.17±0.06	Used	0.81	95.56	0.119	14.20%	Used
C5:1	Tiglylcarnitine	22.9	1.8		0.05 ± 0.01	Excluded	15.64	0.13		0.05±0.01	Excluded	0.37	0.75	0.028	26.10%	Excluded
C5:1-DC	Glutaconylcarnitine	40	24.83		0.01 ± 0.01	Excluded	23.07	13.57		0.02±0	Excluded	0.13	12.48	0.016	42.40%	Excluded
C5-DC (C6-OH)	Glutarylcarnitine (Hydroxyhexanoylcarnitine)	29.4	61.36		0.02 ± 0.01	Excluded	21.23	47.29		0.02±0.01	Excluded	0.15	27.06	0.041	21.00%	Excluded
C5-M-DC	Methylglutarylcarnitine	28	2.48		0.03 ± 0.01	Excluded	14.42	4.52		0.05±0.05	Excluded	0.18	0.95	0.033	42.90%	Excluded
C5-OH (C3-DC-M)	Hydroxyvalerylcarnitine (Methylmalonylcarnitine)	26.9	19.69		0.03 ± 0.01	Excluded	25.08	18.48		0.04±0.01	Excluded	0.25	55.1	0.027	28.70%	Excluded
C6(C4:1-DC)	Hexanoylcarnitine (Fumarylcarnitine)	21.8	65.33		0.09 ± 0.08	Used	21.65	43.28		0.07±0.04	Excluded	0.85	76.67	0.073	13.60%	Used
C6:1	Hexenoylcarnitine	30.7	5.2		0.02 ± 0.01	Excluded	23.52	5.81		0.02±0.01	Excluded	0.07	0.33	0.018	32.40%	Excluded
C7-DC	Pimelylcarnitine	18.4	70.53		0.05 ± 0.02	Used	13.73	63.7		0.03±0.01	Used	0.79	61.34	0.045	34.40%	Excluded
C8	Octanoylcarnitine	13.2	60.62		0.27 ± 0.24	Used	9.06	46.9		0.14±0.09	Excluded	0.89	51.54	0.223	16.30%	Used
C8:1	Octenoylcarnitine											0.92	96.01	0.09	8.40%	Used
C9	Nonaylcarnitine	23.6	97.28		0.05 ± 0.02	Used	21.26	94.7		0.05±0.02	Excluded	0.84	83.73	0.05	20.80%	Used
C10	Decanoylcarnitine	11.7	99.07		0.39 ± 0.3	Used	8.52	89.15		0.23±0.13	Used	0.93	94.08	0.36	11.40%	Used
C10:1	Decenoylcarnitine	11.2	74.8		0.18 ± 0.08	Used	7.91	46.12		0.13±0.05	Excluded	0.83	48.66	0.169	10.40%	Used
C10:2	Decadienylcarnitine	16	94.86		0.04 ± 0.01	Used	10.32	88.37		0.04±0.02	Used	0.51	50.49	0.042	14.50%	Used
C12	Dodecanoylcarnitine	12.2	96.41		0.16 ± 0.07	Used	9.85	58.14		0.1±0.04	Used	0.86	87.35	0.131	10.40%	Used
C12:1	Dodecenoylcarnitine	15.2	26.75		0.17 ± 0.06	Excluded	10.08	38.24		0.09±0.04	Excluded	0.73	13.69	0.147	13.00%	Used
C12-DC	Dodecanedioylcarnitine	12.3	0		0.07 ± 0.01	Excluded	10.25	0		0.07±0.01	Excluded	0.05	0	0.058	12.20%	Excluded
C14	Tetradecanoylcarnitine	15.8	96.66		0.06 ± 0.02	Used	12.59	85.4		0.05±0.01	Used	0.54	51.67	0.046	12.60%	Used

C14:2	Tetradecadienylcarnitine	18.3	98.33	0.04 ± 0.02	Used	12.29	95.87	0.02±0.01	Used	0.87	98.82	0.032	11.60 %	Used
C14:2-OH	Hydroxytetradecadienylcarnitine	35.1	47	0.01 ± 0	Excluded	16.17	35.01	0.01±0.01	Excluded	0.27	38.04	0.009	17.40 %	Excluded
C16	Hexadecanylcarnitine	11.3	99.63	0.14 ± 0.03	Used	10.48	100	0.13±0.03	Used	0.84	100	0.118	8.90%	Used
C16-OH	Hydroxyhexadecanoylcarnitine	33	16.28	0.01 ± 0	Excluded	22.03	51.03	0.01±0	Excluded	0.2	3.33	0.007	24.10 %	Excluded
C16:1	Hexadecenoylcarnitine	18.1	77.83	0.04 ± 0.01	Used	13.83	61.37	0.03±0.01	Used	0.71	2.78	0.038	10.20 %	Used
C16:1-OH	Hydroxyhexadecenoylcarnitine	26.5	26.01	0.01 ± 0	Excluded	24.69	45.09	0.01±0	Excluded	0.38	2.25	0.011	17.50 %	Excluded
C16:2	Hexadecadienylcarnitine	34	87.49	0.01 ± 0	Excluded	16.35	72.35	0.01±0	Used	0.57	70.69	0.008	19.40 %	Used
C16:2-OH	Hydroxyhexadecadienylcarnitine	30.1	5.76	0.01 ± 0	Excluded	21.38	0.13	0.01±0	Excluded	0.32	4.67	0.012	16.60 %	Excluded
C18	Octadecanoylcarnitine	15.7	99.63	0.06 ± 0.01	Used	15.04	100	0.06±0.02	Used	0.69	99.8	0.05	13.70 %	Used
C18:1	Octadecenoylcarnitine	9.7	99.57	0.15 ± 0.04	Used	8.76	100	0.17±0.05	Used	0.87	98.33	0.13	10.20 %	Used
C18:1-OH	Hydroxyoctadecenoylcarnitine	44.6	7.37	0.01 ± 0	Excluded	24.11	2.07	0.01±0	Excluded	0.06	0.95	0.011	33.40 %	Excluded
C18:2	Octadecadienylcarnitine	10.5	99.57	0.05 ± 0.02	Used	8.32	100	0.07±0.02	Used	0.81	100	0.047	9.40%	Used
Ala	Alanine	13.7	99.5	419.93 ± 100.25	Used	5.97	100	474.5±97.72	Used					
Arg	Arginine	13.2	99.26	127.38 ± 27.01	Used	12.26	100	135.03±30.41	Used	0.59	100	116.025	8.20%	Used
Asn	Asparagine	11.1	99.57	46.08 ± 8.47	Used	22.48	100	56.13±12.52	Excluded					
Asp	Aspartate	12.2	99.44	30.05 ± 9.93	Used	18.39	100	51.14±12.92	Used					
Cit	Citrulline	12.7	99.44	35.11 ± 9.94	Used	11.2	100	36.92±11.83	Used					
Gln	Glutamine	12.8	99.57	579.79 ± 118.55	Used	15.13	100	612.01±115.41	Used	0.62	100	621.149	9.90%	Used
Glu	Glutamate	15.8	99.57	81.53 ± 35.4	Used	10.53	100	170.97±75.34	Used					
Gly	Glycine	13.2	99.5	263.02 ± 74.35	Used	16.76	100	347.87±81.28	Used	0.89	100	307.653	7.90%	Used
His	Histidine	12.9	99.38	82.39 ± 15.13	Used	12.47	100	98.59±17.24	Used	0.69	100	98.126	8.30%	Used
Ile	Isoleucine	13.9	99.63	72.65 ± 19.67	Used	12.03	100	98.52±32.16	Used					
Leu	Leucine	12.9	98.58	160.26 ± 43.19	Used	10.41	100	204.78±54.9	Used					
Lys	Lysine	15.5	99.69	166.36 ± 37.05	Used	14.05	100	176.72±34	Used					
Met	Methionine	13.5	99.69	23.94 ± 6.85	Used	13.62	100	26.33±7.76	Used	0.53	100	32.208	9.70%	Used
Orn	Ornithine	14.9	99.63	59.34 ± 15.12	Used	15.48	100	94.12±21.6	Used	0.75	100	81.891	9.40%	Used
Phe	Phenylalanine	12.2	99.57	77.51 ± 17.62	Used	12.11	100	109.69±18.79	Used	0.62	100	62.269	8.40%	Used
Pro	Proline	11.8	99.63	195.17 ± 60.16	Used	9.23	99.87	250.07±63.9	Used	0.89	100	176.445	7.40%	Used
Ser	Serine	13.6	99.44	126.75 ± 28.92	Used	13.6	100	161.02±29.55	Used	0.62	100	128.342	9.60%	Used
Thr	Threonine	18.3	99.13	120.31 ± 31.71	Used	5.55	100	123.6±28.35	Used	0.71	100	105.941	12.10 %	Used
Trp	Tryptophan	12.9	99.63	60.23 ± 12.64	Used	11.5	100	68.65±14.31	Used	0.51	100	82.656	7.50%	Used
Tyr	Tryosine	14.7	99.57	72.51 ± 20.2	Used	11.62	100	82.62±21.04	Used	0.66	100	85.554	8.60%	Used
Val	Valine	13.5	99.63	228.02 ± 51.89	Used	12	100	243.31±50.16	Used	0.69	100	278.405	19.60 %	Used
Ac-Orn	Acetylornithine	20.8	79.07	0.75 ± 0.46	Used	37.51	96.38	1.35±1.08	Excluded	0.74	100	214.254	8.20%	Used
ADMA	Asymmetric dimethylarginine	17.4	66.5	0.57 ± 0.38	Used	26.42	100	0.49±0.13	Excluded					
SDMA	Symmetric dimethylarginine	32.4	97.34	0.77 ± 0.26	Excluded	41.52	99.22	0.56±0.28	Excluded					
total-DMA	Total dimethylarginine	20.3	99.2	1.22 ± 0.25	Used	25.49	89.92	0.71±0.34	Excluded					
alpha-AAA	alpha-Aminoadipic acid	32	97.34	0.67 ± 0.29	Excluded	26.89	95.87	1.03±0.67	Excluded					
Carnosine	Carnosine	89.8	4.02	0.13 ± 0.06	Excluded	NA	0.78	0.07±0.1	Excluded					
Creatinine	Creatinine	14.7	99.38	76.07 ± 24.21	Used	5.79	100	80.44±44.11	Used					
Histamine	Histamine	43.5	89.97	0.39 ± 0.17	Excluded	77.23	53.62	0.15±0.16	Excluded					
Kynurenine	Kynurenine	11.3	97.28	2.93 ± 0.76	Used	21.59	99.87	2.67±0.91	Excluded					
Met-SO	Methioninesulfoxide	20.9	96.66	0.76 ± 0.27	Used	20.55	99.22	1.22±0.6	Excluded					
Nitro-Tyr	Nitrotyrosine	58.4	7.55	0.66 ± 0.32	Excluded	NA	0.39	0.35±0.37	Excluded					
OH-Pro	Hydroxyproline	NA	2.11	5.34 ± 5.15	Excluded	165.29	2.07	1.67±6.16	Excluded					
PEA	Phenylethylamine	NA	0.56	0.09 ± 0.01	Excluded	20.14	0.52	0.02±0.04	Excluded					

Putrescine	Putrescine	53.2	93.75	0.15 ± 0.05	Excluded	56.69	69.38	0.06±0.26	Excluded					
Sarcosine	Sarcosine	28.7	4.4	93.9 ± 223.85	Excluded	83.02	99.61	11.52±14.13	Excluded					
Serotonin	Serotonin	38	99.32	0.69 ± 0.36	Excluded	131.4	99.22	0.57±0.29	Excluded					
Spermidine	Spermidine	24.1	98.51	0.27 ± 0.07	Used	19.46	74.03	0.23±0.27	Used					
Spermine	Spermine	8.5	9.29	0.28 ± 0.13	Excluded	19.74	63.82	0.28±0.3	Used					
Taurine	Taurine	13.7	96.9	93.65 ± 24.37	Used	11.42	100	141.64±22.76	Used					
DOPA	DOPA	19.5	44.58	0.49 ± 0.58	Excluded	47.12	20.67	0.2±0.11	Excluded					
Dopamine	Dopamine	NA	0.06	0.61 ± NA	Excluded	77.04	1.42	0.11±2.12	Excluded					
lysoPC a C6:0	lysoPhosphatidylcholine acyl C6:0									-0.14	33.33	0.023	62.50%	Excluded
lysoPC a C14:0	lysoPhosphatidylcholine acyl C14:0	6.8	0	6.28 ± 0.8	Excluded	5.45	42.38	5.82±1.53	Excluded	0.45	21.24	3.209	23.80%	Excluded
lysoPC a C16:0	lysoPhosphatidylcholine acyl C16:0	6.9	99.81	122.71 ± 26.13	Used	6.75	100	233.73±55.43	Used	0.75	100	94.408	8.80%	Used
lysoPC a C16:1	lysoPhosphatidylcholine acyl C16:1	7	99.69	3.71 ± 1.25	Used	7.59	100	6.73±2.16	Used	0.84	100	2.909	8.60%	Used
lysoPC a C17:0	lysoPhosphatidylcholine acyl C17:0	7.3	99.63	2.08 ± 0.67	Used	7.54	100	4.58±1.67	Used	0.84	100	1.727	12.70%	Used
lysoPC a C18:0	lysoPhosphatidylcholine acyl C18:0	7.2	99.81	32.73 ± 8.11	Used	8.01	100	64.98±18.47	Used	0.8	100	26.066	9.70%	Used
lysoPC a C18:1	lysoPhosphatidylcholine acyl C18:1	6.8	99.75	21.5 ± 6.28	Used	7.65	100	35.79±9.82	Used	0.84	100	19.277	9.20%	Used
lysoPC a C18:2	lysoPhosphatidylcholine acyl C18:2	6.9	99.75	28.09 ± 8.99	Used	7.18	100	46.06±15.11	Used	0.93	100	27.241	8.80%	Used
lysoPC a C20:3	lysoPhosphatidylcholine acyl C20:3	8.8	99.63	2.29 ± 0.68	Used	7.41	100	3.63±1.11	Used	0.77	100	2.393	9.00%	Used
lysoPC a C20:4	lysoPhosphatidylcholine acyl C20:4	7.3	99.69	6.2 ± 1.88	Used	7.32	100	10.38±3.03	Used	0.87	100	6.798	9.00%	Used
lysoPC a C24:0	lysoPhosphatidylcholine acyl C24:0	32	23.22	0.21 ± 0.07	Excluded	23.98	64.47	0.32±0.11	Excluded	0.09	12.45	0.364	21.10%	Excluded
lysoPC a C26:0	lysoPhosphatidylcholine acyl C26:0	44.4	43.72	0.3 ± 0.16	Excluded	45.37	98.06	0.55±0.31	Excluded	0.09	59.58	0.548	31.00%	Excluded
lysoPC a C26:1	lysoPhosphatidylcholine acyl C26:1	9.5	0	1.68 ± 0.17	Excluded	39.74	99.87	0.25±0.12	Excluded	-0.04	0	2.027	7.90%	Excluded
lysoPC a C28:0	lysoPhosphatidylcholine acyl C28:0	37	23.47	0.33 ± 0.12	Excluded	34.67	95.35	0.55±0.23	Excluded	0.17	49.61	0.492	29.10%	Excluded
lysoPC a C28:1	lysoPhosphatidylcholine acyl C28:1	35.5	98.64	0.47 ± 0.16	Excluded	29.45	100	0.69±0.23	Excluded	0.29	99.84	0.618	22.60%	Excluded
PC aa C24:0	Phosphatidylcholine diacyl C24:0	45.9	69.35	0.09 ± 0.05	Excluded	39.09	96.51	0.16±0.08	Excluded	0.11	72.55	0.15	26.50%	Excluded
PC aa C26:0	Phosphatidylcholine diacyl C26:0	27.2	5.63	0.74 ± 0.26	Excluded	27.32	29.46	1.16±0.41	Excluded	0.09	11.54	1.05	32.90%	Excluded
PC aa C28:1	Phosphatidylcholine diacyl C28:1	9.5	99.63	3.58 ± 0.84	Used	7.65	100	3.22±0.82	Used	0.87	100	3.362	9.80%	Used
PC aa C30:0	Phosphatidylcholine diacyl C30:0	9.4	99.63	5.85 ± 1.71	Used	7.28	100	4.59±1.43	Used	0.89	100	4.716	7.80%	Used
PC aa C30:2	Phosphatidylcholine diacyl C30:2	89.9	31.33	0.03 ± 0.07	Excluded	51.75	89.28	0.12±0.09	Excluded	0.12	4.22	0.003	81.60%	Excluded
PC aa C32:0	Phosphatidylcholine diacyl C32:0	8.4	99.81	14.86 ± 2.85	Used	5.91	100	14.16±3.08	Used	0.83	100	15.145	7.10%	Used
PC aa C32:1	Phosphatidylcholine diacyl C32:1	9.2	99.81	20.81 ± 10.63	Used	7.95	100	15.37±10.01	Used	0.96	100	21.842	7.40%	Used
PC aa C32:2	Phosphatidylcholine diacyl C32:2	12.3	99.81	4.31 ± 1.66	Used	8.64	100	4.42±1.65	Used	0.91	99.93	3.887	11.10%	Used
PC aa C32:3	Phosphatidylcholine diacyl C32:3	9.2	99.75	0.54 ± 0.14	Used	9.25	100	0.6±0.18	Used	0.79	100	0.481	8.90%	Used
PC aa C34:1	Phosphatidylcholine diacyl C34:1	7.1	99.88	222.04 ± 46.61	Used	5.78	100	205.3±53.16	Used	0.83	100	240.267	7.20%	Used
PC aa C34:2	Phosphatidylcholine diacyl C34:2	7	99.88	364.87 ± 50.66	Used	5.86	100	351.39±63.18	Used	0.75	100	391.393	7.70%	Used
PC aa C34:3	Phosphatidylcholine diacyl C34:3	6.3	99.88	18.18 ± 5.25	Used	6.51	100	16.73±4.97	Used	0.91	100	17.892	8.60%	Used
PC aa C34:4	Phosphatidylcholine diacyl C34:4	6.8	99.81	2.18 ± 0.77	Used	7.43	100	2.24±0.76	Used	0.92	100	2.261	8.00%	Used
PC aa C36:0	Phosphatidylcholine diacyl C36:0	11.6	99.63	2.98 ± 0.75	Used	9.15	100	2.63±0.72	Used	0.74	100	2.69	17.40%	Used
PC aa C36:1	Phosphatidylcholine diacyl C36:1	6.9	99.88	53.54 ± 14.16	Used	5.58	100	55.79±14.67	Used	0.84	100	53.831	8.50%	Used
PC aa C36:2	Phosphatidylcholine diacyl C36:2	6.6	99.88	255.06 ± 43.56	Used	5.25	100	241.28±44.22	Used	0.8	100	231.934	6.70%	Used
PC aa C36:3	Phosphatidylcholine diacyl C36:3	6.5	99.88	151.97 ± 29.94	Used	5.97	100	155.15±31.97	Used	0.86	100	149.595	7.50%	Used
PC aa C36:4	Phosphatidylcholine diacyl C36:4	6.3	99.94	208.67 ± 44.62	Used	5.86	100	201.24±44.34	Used	0.87	100	219.578	7.80%	Used
PC aa C36:5	Phosphatidylcholine diacyl C36:5	6.7	99.81	30.78 ± 15.23	Used	6.11	100	29.18±13.97	Used	0.82	100	29.302	8.60%	Used
PC aa C36:6	Phosphatidylcholine diacyl C36:6	9.5	99.75	1.11 ± 0.43	Used	7.54	100	1.33±0.51	Used	0.89	100	1.118	11.10%	Used
PC aa C38:0	Phosphatidylcholine diacyl C38:0	8.8	99.63	3.35 ± 0.86	Used	6.57	100	3.84±0.92	Used	0.86	100	3.256	13.80%	Used
PC aa C38:1	Phosphatidylcholine diacyl C38:1	27	99.75	1.38 ± 0.43	Excluded	17.47	100	1.58±0.42	Used	0.34	99.84	0.859	18.10%	Excluded

PC aa C38:3	Phosphatidylcholine diacyl C38:3	6.9	99.88	58.03 ± 14.11	Used	5.02	100	54.98±13.85	Used	0.86	100	53.993	7.60%	Used
PC aa C38:4	Phosphatidylcholine diacyl C38:4	5.7	99.88	119.01 ± 29.37	Used	5.26	100	111.23±26.31	Used	0.88	100	119.655	7.30%	Used
PC aa C38:5	Phosphatidylcholine diacyl C38:5	5.6	99.88	62.89 ± 15.21	Used	5.54	100	56.45±13.65	Used	0.83	100	62.173	7.90%	Used
PC aa C38:6	Phosphatidylcholine diacyl C38:6	6.9	100	89.9 ± 26.07	Used	5.85	100	92.7±24.75	Used	0.93	100	90.24	8.10%	Used
PC aa C40:1	Phosphatidylcholine diacyl C40:1	11.7	14.24	0.42 ± 0.09	Excluded	9.39	11.24	0.5±0.11	Excluded	0.51	8.66	0.465	13.50 %	Used
PC aa C40:2	Phosphatidylcholine diacyl C40:2	14.9	99.63	0.37 ± 0.11	Used	10.52	100	0.48±0.14	Used	0.51	100	0.357	11.70 %	Used
PC aa C40:3	Phosphatidylcholine diacyl C40:3	13.9	99.75	0.69 ± 0.16	Used	7.6	100	0.82±0.21	Used	0.6	100	0.653	11.20 %	Used
PC aa C40:4	Phosphatidylcholine diacyl C40:4	6.8	99.81	4.15 ± 1.22	Used	5.58	100	4.2±1.2	Used	0.86	100	4.157	7.60%	Used
PC aa C40:5	Phosphatidylcholine diacyl C40:5	6.5	99.75	12.75 ± 3.6	Used	5.23	100	13.43±4.04	Used	0.89	100	11.504	7.00%	Used
PC aa C40:6	Phosphatidylcholine diacyl C40:6	6.1	99.63	32.47 ± 10.04	Used	5.48	100	34.36±10.66	Used	0.93	100	28.751	7.10%	Used
PC aa C42:0	Phosphatidylcholine diacyl C42:0	9.2	99.88	0.55 ± 0.15	Used	8.3	100	0.52±0.15	Used	0.85	99.97	0.592	12.30 %	Used
PC aa C42:1	Phosphatidylcholine diacyl C42:1	12	99.69	0.27 ± 0.07	Used	9.76	100	0.26±0.07	Used	0.72	100	0.297	14.80 %	Used
PC aa C42:2	Phosphatidylcholine diacyl C42:2	13.5	99.69	0.2 ± 0.05	Used	9.13	100	0.24±0.06	Used	0.56	100	0.209	14.60 %	Used
PC aa C42:4	Phosphatidylcholine diacyl C42:4	11	99.81	0.21 ± 0.04	Used	10.46	100	0.17±0.04	Used	0.51	100	0.219	11.70 %	Used
PC aa C42:5	Phosphatidylcholine diacyl C42:5	11.3	99.69	0.44 ± 0.13	Used	8.32	100	0.51±0.15	Used	0.75	100	0.426	10.60 %	Used
PC aa C42:6	Phosphatidylcholine diacyl C42:6	10.7	95.42	0.6 ± 0.14	Used	8.39	69.51	0.72±0.17	Used	0.62	60.16	0.626	12.50 %	Used
PC ae C30:0	Phosphatidylcholine acyl-alkyl C30:0	19.7	99.57	0.45 ± 0.13	Used	14.01	97.93	0.41±0.12	Used	0.76	98.86	0.464	18.10 %	Used
PC ae C30:1	Phosphatidylcholine acyl-alkyl C30:1	77.9	82.35	0.13 ± 0.11	Excluded	53.44	99.74	0.23±0.11	Excluded	0.18	94.12	0.224	41.70 %	Excluded
PC ae C30:2	Phosphatidylcholine acyl-alkyl C30:2	25.2	99.57	0.13 ± 0.04	Excluded	16.46	95.87	0.17±0.04	Used	0.65	86.34	0.156	17.50 %	Used
PC ae C32:1	Phosphatidylcholine acyl-alkyl C32:1	9.3	99.81	2.79 ± 0.56	Used	6.82	100	2.7±0.57	Used	0.83	100	2.852	8.00%	Used
PC ae C32:2	Phosphatidylcholine acyl-alkyl C32:2	12.2	99.63	0.71 ± 0.16	Used	9.66	100	0.78±0.18	Used	0.77	100	0.748	11.60 %	Used
PC ae C34:0	Phosphatidylcholine acyl-alkyl C34:0	9.6	99.81	1.67 ± 0.43	Used	6.84	100	1.85±0.49	Used	0.82	100	1.724	7.90%	Used
PC ae C34:1	Phosphatidylcholine acyl-alkyl C34:1	7.4	99.81	10.4 ± 2.19	Used	6.29	100	9±2.07	Used	0.87	100	10.516	7.60%	Used
PC ae C34:2	Phosphatidylcholine acyl-alkyl C34:2	7.2	99.88	11.96 ± 2.96	Used	6.33	100	11.07±2.69	Used	0.9	100	12.608	7.60%	Used
PC ae C34:3	Phosphatidylcholine acyl-alkyl C34:3	6.9	99.88	7.51 ± 2.12	Used	6.22	100	8.01±2.32	Used	0.91	100	8.34	7.90%	Used
PC ae C36:0	Phosphatidylcholine acyl-alkyl C36:0	22.7	99.63	0.93 ± 0.28	Used	12.83	100	1.08±0.27	Used	0.35	100	1.062	35.60 %	Excluded
PC ae C36:1	Phosphatidylcholine acyl-alkyl C36:1	7.9	99.75	8.79 ± 2.05	Used	6.25	100	8.2±1.99	Used	0.85	100	8.303	9.80%	Used
PC ae C36:2	Phosphatidylcholine acyl-alkyl C36:2	7	99.88	15.06 ± 3.84	Used	5.84	100	15.84±4.33	Used	0.92	100	15.064	8.30%	Used
PC ae C36:3	Phosphatidylcholine acyl-alkyl C36:3	7.1	99.88	7.91 ± 1.85	Used	6.37	100	8.54±1.84	Used	0.86	100	8.545	8.10%	Used
PC ae C36:4	Phosphatidylcholine acyl-alkyl C36:4	6.3	99.88	19.99 ± 4.78	Used	6.33	100	17.87±4.09	Used	0.87	100	20.774	7.90%	Used
PC ae C36:5	Phosphatidylcholine acyl-alkyl C36:5	6.1	99.81	12.96 ± 3.22	Used	6.28	100	13.01±3.29	Used	0.89	100	13.807	8.00%	Used
PC ae C38:0	Phosphatidylcholine acyl-alkyl C38:0	8.1	99.63	2.22 ± 0.66	Used	6.18	100	2.76±0.78	Used	0.81	100	2.465	10.80 %	Used
PC ae C38:1	Phosphatidylcholine acyl-alkyl C38:1	14.7	99.5	0.62 ± 0.26	Used	11.96	34.75	0.09±0.17	Excluded	0.48	100	0.814	12.40 %	Used
PC ae C38:2	Phosphatidylcholine acyl-alkyl C38:2	11.7	99.75	2.1 ± 0.48	Used	7.05	100	2.48±0.57	Used	0.73	100	2.123	10.30 %	Used
PC ae C38:3	Phosphatidylcholine acyl-alkyl C38:3	7	99.94	4.25 ± 0.98	Used	6.04	100	5.1±1.2	Used	0.85	100	4.295	9.20%	Used
PC ae C38:4	Phosphatidylcholine acyl-alkyl C38:4	6.1	100	15.36 ± 3.11	Used	5.73	100	14.36±2.77	Used	0.82	100	15.628	8.60%	Used
PC ae C38:5	Phosphatidylcholine acyl-alkyl C38:5	5.9	100	19.25 ± 3.86	Used	5.76	100	19.64±3.74	Used	0.82	100	19.863	8.30%	Used
PC ae C38:6	Phosphatidylcholine acyl-alkyl C38:6	6.5	99.88	8.7 ± 2.05	Used	5.93	100	8.57±1.91	Used	0.85	100	8.677	8.10%	Used
PC ae C40:0	Phosphatidylcholine acyl-alkyl C40:0									0.87	1.05	10.205	4.80%	Used
PC ae C40:1	Phosphatidylcholine acyl-alkyl C40:1	11.1	99.63	1.58 ± 0.37	Used	9.42	100	1.83±0.41	Used	0.68	100	1.661	10.50 %	Used
PC ae C40:2	Phosphatidylcholine acyl-alkyl C40:2	8.3	99.88	2.12 ± 0.49	Used	6.2	100	2.17±0.53	Used	0.85	100	2.083	9.50%	Used

PC ae C40:3	Phosphatidylcholine acyl-alkyl C40:3	9	99.94	1.18 ± 0.24	Used	6.29	100	1.18±0.32	Used	0.73	100	1.133	9.50%	Used
PC ae C40:4	Phosphatidylcholine acyl-alkyl C40:4	8.7	99.63	2.68 ± 0.51	Used	6.57	100	2.83±0.56	Used	0.82	100	2.568	9.60%	Used
PC ae C40:5	Phosphatidylcholine acyl-alkyl C40:5	6.5	99.88	3.66 ± 0.69	Used	5.72	100	4.4±0.88	Used	0.78	100	3.547	8.30%	Used
PC ae C40:6	Phosphatidylcholine acyl-alkyl C40:6	6.9	99.94	5.42 ± 1.3	Used	5.62	100	5.7±1.35	Used	0.88	100	5.04	8.60%	Used
PC ae C42:0	Phosphatidylcholine acyl-alkyl C42:0	13.8	36.35	0.52 ± 0.1	Excluded	8.46	15.89	0.69±0.12	Excluded	0.6	14.87	0.513	15.70 %	Used
PC ae C42:1	Phosphatidylcholine acyl-alkyl C42:1	16	99.57	0.38 ± 0.09	Used	13.43	100	0.48±0.11	Used	0.51	100	0.371	11.50 %	Used
PC ae C42:2	Phosphatidylcholine acyl-alkyl C42:2	11.5	99.69	0.63 ± 0.14	Used	8.13	100	0.72±0.16	Used	0.69	100	0.664	12.80 %	Used
PC ae C42:3	Phosphatidylcholine acyl-alkyl C42:3	9.8	99.88	0.84 ± 0.19	Used	8.46	100	0.86±0.18	Used	0.8	100	0.861	10.80 %	Used
PC ae C42:4	Phosphatidylcholine acyl-alkyl C42:4	7.8	99.63	0.95 ± 0.22	Used	6.21	100	0.96±0.22	Used	0.78	100	1.007	9.20%	Used
PC ae C42:5	Phosphatidylcholine acyl-alkyl C42:5	7.4	99.57	2.19 ± 0.46	Used	5.55	100	2.26±0.49	Used	0.86	99.97	2.346	7.40%	Used
PC ae C44:3	Phosphatidylcholine acyl-alkyl C44:3	24.3	99.69	0.13 ± 0.04	Used	14.93	100	0.14±0.03	Used	0.5	100	0.11	12.50 %	Used
PC ae C44:4	Phosphatidylcholine acyl-alkyl C44:4	12.1	99.69	0.38 ± 0.1	Used	9.21	100	0.36±0.08	Used	0.71	100	0.427	11.40 %	Used
PC ae C44:5	Phosphatidylcholine acyl-alkyl C44:5	7.4	99.69	1.73 ± 0.46	Used	6.18	100	1.46±0.38	Used	0.86	100	2.112	8.00%	Used
PC ae C44:6	Phosphatidylcholine acyl-alkyl C44:6	7.8	99.63	1.25 ± 0.33	Used	6.16	100	1.02±0.28	Used	0.89	100	1.373	7.70%	Used
SM (OH) C14:1	Hydroxysphingomyeline C14:1	11	99.63	9.43 ± 2.53	Used	7.18	100	9.88±2.94	Used	0.91	100	6.222	7.70%	Used
SM (OH) C16:1	Hydroxysphingomyeline C16:1	11	100	5.15 ± 1.35	Used	8.16	100	4.85±1.39	Used	0.86	100	3.353	8.80%	Used
SM (OH) C22:1	Hydroxysphingomyeline C22:1	11.2	99.88	20.18 ± 4.56	Used	8.63	100	18.97±4.74	Used	0.82	100	13.496	11.20 %	Used
SM (OH) C22:2	Hydroxysphingomyeline C22:2	11.2	99.88	16.33 ± 4.09	Used	8.94	100	14.95±4.16	Used	0.87	100	11.409	10.30 %	Used
SM (OH) C24:1	Hydroxysphingomyeline C24:1	15.1	99.75	1.99 ± 0.49	Used	10.86	100	1.76±0.48	Used	0.75	100	1.344	15.10 %	Used
SM C16:0	Sphingomyeline C16:0	10.6	99.88	150.28 ± 24.31	Used	7.28	100	146.17±24.55	Used	0.73	100	106.649	8.00%	Used
SM C16:1	Sphingomyeline C16:1	9.9	99.88	23.81 ± 4.67	Used	7.33	100	23.75±5.19	Used	0.84	100	16.059	7.50%	Used
SM C18:0	Sphingomyeline C18:0	9.8	99.81	33.18 ± 7.02	Used	7.76	100	31.52±7.17	Used	0.79	100	23.224	9.00%	Used
SM C18:1	Sphingomyeline C18:1	9.4	99.88	16.7 ± 4.09	Used	7.3	100	15.29±4.12	Used	0.84	100	11.304	8.20%	Used
SM C20:2	Sphingomyeline C20:2	16.2	99.81	0.66 ± 0.23	Used	9.59	100	1.21±0.62	Used	0.61	99.93	0.381	12.60 %	Used
SM C22:3	Sphingomyeline C22:3	NA	0.37	0 ± 0.01	Excluded	41.07	78.94	0.39±0.41	Excluded	-0.04	55.85	0.114	57.60 %	Excluded
SM C24:0	Sphingomyeline C24:0	11.9	99.75	30.26 ± 5.82	Used	8.81	100	28.77±5.78	Used	0.78	100	21.79	10.70 %	Used
SM C24:1	Sphingomyeline C24:1	12.1	99.88	76.6 ± 14.93	Used	8.79	100	66.12±13.01	Used	0.75	100	52.568	10.00 %	Used
SM C26:0	Sphingomyeline C26:0	31.8	99.81	0.3 ± 0.09	Excluded	32.47	100	0.21±0.07	Excluded	0.46	100	0.176	67.80 %	Excluded
SM C26:1	Sphingomyeline C26:1	21.2	99.75	0.65 ± 0.2	Used	17.62	100	0.57±0.17	Used	0.69	100	0.417	20.80 %	Used
H1	Hexose	5.2	99.81	5368.96 ± 1010.34	Used	5.53	100	5255.8±1967.33	Used	0.69	100	5190.295	6.30%	Used

Table S2 CpG sites different between S and never smokers in KORA F4 and replicated in KORA F3

		F4		F3	
CpG site	**Gene**	**Association β (95% CI)**	**P**	**Association β (95% CI)**	**P**
cg22635096	ADARB1	0.22(0.15,0.29)	1.98E-10	0.19(0.12,0.25)	3.43E-08
cg01207684	ADCY9	-0.25(-0.31,-0.20)	3.42E-22	-0.23(-0.30,-0.17)	7.03E-12
cg05765011	ADCY9	-0.17(-0.22,-0.13)	7.98E-13	-0.11(-0.16,-0.06)	1.48E-05
cg00689360	ADH1C	-0.28(-0.38,-0.18)	6.16E-08	-0.22(-0.33,-0.11)	6.75E-05
cg05329352	ADRA2A	-0.15(-0.19,-0.11)	3.21E-14	-0.12(-0.17,-0.07)	1.13E-06
cg05575921	AHRR	-1.76(-1.85,-1.66)	1.26E-215	-1.84(-2.00,-1.68)	4.67E-76
cg26703534	AHRR	-0.40(-0.43,-0.37)	2.39E-123	-0.35(-0.40,-0.31)	3.54E-41
cg21161138	AHRR	-0.67(-0.73,-0.62)	1.37E-113	-0.64(-0.72,-0.56)	5.68E-47
cg25648203	AHRR	-0.75(-0.82,-0.68)	6.54E-89	-0.66(-0.76,-0.56)	2.72E-34
cg14817490	AHRR	-0.56(-0.63,-0.50)	1.69E-59	-0.43(-0.50,-0.36)	3.86E-27
cg03991871	AHRR	-1.15(-1.30,-1.00)	7.14E-50	-1.18(-1.37,-0.99)	2.12E-29
cg11902777	AHRR	-0.51(-0.58,-0.44)	1.24E-44	-0.39(-0.47,-0.32)	6.41E-23
cg23576855	AHRR	-1.46(-1.66,-1.26)	8.66E-44	-1.39(-1.62,-1.16)	9.29E-29
cg01899089	AHRR	-0.28(-0.32,-0.24)	1.02E-41	-0.19(-0.23,-0.14)	1.57E-17
cg12806681	AHRR	-0.72(-0.82,-0.62)	1.90E-41	-0.64(-0.77,-0.51)	5.63E-21
cg11554391	AHRR	-0.23(-0.27,-0.20)	3.66E-33	-0.15(-0.19,-0.12)	7.00E-16
cg23916896	AHRR	-0.31(-0.37,-0.26)	1.48E-31	-0.27(-0.32,-0.22)	1.15E-22
cg03604011	AHRR	0.51(0.42,0.59)	4.24E-28	0.30(0.20,0.40)	7.08E-09
cg24090911	AHRR	-0.42(-0.51,-0.33)	3.61E-19	-0.35(-0.47,-0.24)	5.41E-09
cg17924476	AHRR	0.31(0.24,0.39)	1.52E-17	0.21(0.14,0.28)	2.34E-08
cg10841124	AHRR	0.26(0.20,0.32)	2.46E-17	0.19(0.11,0.26)	2.47E-06
cg16219322	AHRR	-0.26(-0.33,-0.19)	4.84E-13	-0.19(-0.25,-0.12)	7.52E-08
cg26850624	AHRR	0.18(0.13,0.24)	4.36E-12	0.17(0.10,0.23)	9.20E-07
cg09338136	AHRR	-0.06(-0.08,-0.04)	1.88E-08	-0.09(-0.12,-0.06)	1.40E-08
cg24688690	AHRR	-0.11(-0.14,-0.07)	2.06E-08	-0.09(-0.13,-0.05)	2.31E-06
cg22356527	AHRR	0.10(0.07,0.14)	4.16E-08	0.08(0.05,0.11)	2.16E-08
cg11557553	AHRR	0.20(0.14,0.26)	3.45E-10	0.14(0.08,0.20)	1.98E-05
cg22937882	AHRR	0.19(0.12,0.26)	9.56E-08	0.18(0.10,0.26)	2.81E-05
cg03004371	ALDH1A3	0.18(0.12,0.23)	1.21E-09	0.13(0.08,0.19)	4.91E-06
cg08529529	ALOX5AP	-0.12(-0.15,-0.08)	1.46E-11	-0.07(-0.10,-0.04)	3.74E-05
cg19713851	ALPP	-0.34(-0.43,-0.26)	3.62E-14	-0.29(-0.40,-0.18)	1.98E-07
cg23667432	ALPP	-0.10(-0.12,-0.07)	9.20E-12	-0.07(-0.11,-0.04)	2.49E-05
cg03188382	ALPP	-0.12(-0.15,-0.08)	4.42E-11	-0.10(-0.14,-0.06)	2.35E-06
cg22403782	ALPP	-0.13(-0.18,-0.09)	2.18E-08	-0.10(-0.15,-0.05)	7.50E-05
cg03234777	AMICA1	-0.20(-0.25,-0.15)	5.06E-16	-0.11(-0.15,-0.06)	7.90E-06
cg08772028	AMICA1	-0.21(-0.28,-0.15)	1.10E-10	-0.15(-0.22,-0.09)	2.85E-06
cg23161492	ANPEP	-0.26(-0.30,-0.22)	4.33E-37	-0.21(-0.25,-0.16)	2.91E-16
cg06635952	ANXA4	0.18(0.14,0.23)	1.42E-17	0.17(0.12,0.21)	6.84E-11
cg19847577	APBA2	0.17(0.11,0.23)	8.14E-08	0.13(0.07,0.19)	2.26E-05
cg25953130	ARID5B	-0.31(-0.41,-0.21)	1.61E-09	-0.40(-0.51,-0.29)	1.92E-11
cg02186444	ARMC7	0.16(0.12,0.21)	8.63E-14	0.12(0.08,0.17)	2.83E-08

cg01901332	ARRB1	-0.37(-0.44,-0.30)	3.63E-25	-0.32(-0.39,-0.25)	1.31E-18
cg20902353	ASB2	-0.11(-0.14,-0.08)	2.64E-10	-0.07(-0.10,-0.04)	3.57E-05
cg05478824	ASPSCR1	-0.19(-0.26,-0.12)	5.11E-08	-0.20(-0.28,-0.11)	6.04E-06
cg18405341	ATF4	-0.12(-0.16,-0.08)	2.45E-08	-0.13(-0.18,-0.07)	1.58E-05
cg26337070	ATOH8	-0.25(-0.32,-0.18)	1.33E-12	-0.24(-0.32,-0.16)	1.03E-08
cg00893603	ATP8A2	-0.36(-0.49,-0.24)	4.99E-09	-0.33(-0.47,-0.19)	5.64E-06
cg07339236	ATP9A	-0.25(-0.30,-0.20)	9.74E-24	-0.20(-0.24,-0.16)	2.83E-22
cg20295214	AVPR1B	-0.28(-0.35,-0.20)	1.93E-13	-0.33(-0.40,-0.25)	1.34E-17
cg08709672	AVPR1B	-0.17(-0.20,-0.14)	8.09E-29	-0.13(-0.16,-0.10)	1.93E-13
cg26470501	BCL3	-0.13(-0.17,-0.08)	8.90E-09	-0.12(-0.17,-0.07)	4.70E-06
cg05957567	BCL7A;BCL7A	0.09(0.06,0.11)	5.33E-11	0.07(0.04,0.10)	8.75E-06
cg15482893	BSDC1	0.13(0.09,0.18)	4.98E-09	0.13(0.08,0.17)	4.32E-08
cg10750182	C10orf105	-0.20(-0.22,-0.17)	3.56E-42	-0.18(-0.22,-0.15)	6.47E-24
cg24996979	C14orf43	-0.15(-0.18,-0.12)	5.75E-24	-0.09(-0.12,-0.06)	2.93E-09
cg22851561	C14orf43	-0.30(-0.36,-0.24)	5.35E-21	-0.25(-0.31,-0.19)	3.39E-14
cg01731783	C14orf43	-0.12(-0.15,-0.09)	5.16E-15	-0.11(-0.14,-0.08)	1.86E-10
cg17861836	C17orf56	-0.22(-0.30,-0.15)	9.43E-09	-0.18(-0.27,-0.10)	1.31E-05
cg14580211	C5orf62	-0.33(-0.39,-0.27)	2.26E-28	-0.33(-0.40,-0.27)	1.22E-23
cg08972170	C7orf41	0.28(0.19,0.37)	4.20E-10	0.27(0.17,0.36)	4.22E-08
cg15417641	CACNA1D	0.58(0.44,0.71)	5.09E-17	0.37(0.23,0.50)	2.69E-07
cg00336149	CACNA1D	0.36(0.27,0.46)	1.11E-13	0.26(0.16,0.37)	3.43E-07
cg21188533	CACNA1D	0.54(0.39,0.70)	4.79E-12	0.41(0.24,0.58)	3.72E-06
cg10520740	CACNA2D4	-0.11(-0.15,-0.07)	9.15E-09	-0.09(-0.13,-0.05)	4.22E-05
cg03575602	CAMK1D	-0.23(-0.30,-0.16)	2.28E-11	-0.21(-0.30,-0.11)	2.10E-05
cg23198793	CAPN3	0.14(0.10,0.19)	9.73E-11	0.11(0.07,0.16)	2.43E-06
cg21446172	CAPN8	-0.29(-0.35,-0.23)	5.25E-20	-0.21(-0.30,-0.13)	1.71E-06
cg19713429	CAPZB	-0.12(-0.15,-0.09)	1.58E-13	-0.08(-0.11,-0.05)	1.30E-06
cg01832549	CAPZB	-0.23(-0.30,-0.17)	7.55E-13	-0.16(-0.23,-0.09)	1.19E-05
cg13500388	CBFB	-0.21(-0.28,-0.15)	7.60E-11	-0.12(-0.18,-0.07)	4.11E-06
cg25197194	CCDC48	0.24(0.16,0.33)	2.71E-08	0.24(0.15,0.33)	2.42E-07
cg20303561	CCDC88C	-0.27(-0.36,-0.18)	7.93E-09	-0.19(-0.27,-0.11)	8.01E-06
cg15474579	CDKN1A	-0.27(-0.34,-0.21)	3.07E-17	-0.18(-0.24,-0.13)	4.42E-11
cg21091547	CDKN1A	-0.31(-0.42,-0.21)	5.20E-09	-0.22(-0.31,-0.12)	6.35E-06
cg26364091	CHADL	-0.08(-0.10,-0.05)	7.00E-09	-0.06(-0.08,-0.03)	5.15E-05
cg03194226	CLEC3B	-0.22(-0.29,-0.15)	7.89E-10	-0.20(-0.28,-0.11)	3.76E-06
cg25949550	CNTNAP2	-0.29(-0.33,-0.25)	1.54E-40	-0.22(-0.26,-0.17)	2.71E-20
cg21322436	CNTNAP2	-0.38(-0.43,-0.32)	3.71E-36	-0.26(-0.33,-0.20)	1.56E-14
cg11207515	CNTNAP2	0.46(0.36,0.56)	2.10E-18	0.37(0.25,0.48)	7.14E-10
cg17372101	CNTNAP2	0.29(0.22,0.37)	8.49E-14	0.26(0.18,0.34)	8.42E-10
cg15159987	CPAMD8	-0.21(-0.24,-0.17)	1.29E-26	-0.19(-0.24,-0.15)	1.02E-15
cg02657160	CPOX	-0.29(-0.38,-0.19)	9.60E-10	-0.28(-0.38,-0.18)	4.06E-08
cg23973524	CRTC1	0.20(0.16,0.24)	6.86E-25	0.16(0.10,0.21)	2.44E-08
cg21473814	CRTC1	0.24(0.19,0.29)	7.96E-18	0.14(0.08,0.19)	6.77E-07
cg00501876	CSRNP1	-0.27(-0.31,-0.23)	1.12E-34	-0.21(-0.27,-0.15)	1.43E-10
cg03540589	CSRNP1	-0.29(-0.38,-0.20)	1.11E-09	-0.18(-0.26,-0.09)	6.93E-05
cg24155190	CSRP1	-0.12(-0.16,-0.08)	4.14E-08	-0.09(-0.13,-0.05)	2.43E-06

cg14099685	CUGBP1	-0.12(-0.15,-0.08)	9.04E-09	-0.10(-0.14,-0.06)	7.15E-07
cg14012925	CUX1	0.10(0.06,0.13)	6.48E-08	0.10(0.06,0.15)	4.84E-06
cg01257799	CXCR5	-0.14(-0.17,-0.10)	1.25E-15	-0.06(-0.09,-0.03)	1.67E-05
cg12678834	CXCR5	-0.18(-0.23,-0.12)	3.26E-10	-0.14(-0.17,-0.10)	6.36E-15
cg24342283	CXCR5	-0.13(-0.17,-0.08)	6.60E-09	-0.10(-0.15,-0.06)	1.55E-05
cg01839993	DDIT4	-0.08(-0.11,-0.06)	4.42E-12	-0.08(-0.11,-0.05)	9.78E-07
cg25560398	ECEL1P2	-0.13(-0.16,-0.10)	3.06E-19	-0.09(-0.14,-0.05)	4.75E-05
cg11152412	EDC3	-0.12(-0.16,-0.08)	5.19E-10	-0.07(-0.10,-0.04)	8.17E-06
cg07019857	EFNA5	-0.16(-0.21,-0.11)	1.76E-10	-0.14(-0.19,-0.09)	6.30E-08
cg06459104	EPB41L3	-0.51(-0.65,-0.37)	1.31E-12	-0.41(-0.55,-0.28)	1.10E-08
cg08149865	EPB49	-0.09(-0.11,-0.07)	7.68E-13	-0.07(-0.10,-0.04)	1.64E-06
cg24931658	EPB49	-0.10(-0.13,-0.07)	2.11E-12	-0.10(-0.13,-0.06)	3.37E-07
cg23110422	ETS2	-0.50(-0.61,-0.40)	1.35E-21	-0.35(-0.46,-0.25)	1.24E-10
cg07986378	ETV6	-0.33(-0.43,-0.23)	9.06E-11	-0.39(-0.49,-0.29)	1.11E-13
cg01442064	EVC	-0.06(-0.08,-0.04)	2.07E-09	-0.05(-0.07,-0.03)	1.30E-05
cg03636183	F2RL3	-0.87(-0.94,-0.81)	2.84E-129	-0.96(-1.07,-0.85)	2.41E-50
cg16116321	FAM124B	-0.21(-0.28,-0.15)	2.03E-11	-0.19(-0.26,-0.12)	1.89E-07
cg13876650	FAM54B	-0.06(-0.08,-0.05)	3.05E-12	-0.06(-0.08,-0.04)	6.83E-07
cg06901890	FNBP1	-0.08(-0.11,-0.05)	2.19E-08	-0.07(-0.10,-0.04)	4.76E-06
cg04517079	FOXP4	-0.16(-0.20,-0.12)	3.00E-14	-0.16(-0.20,-0.12)	1.03E-13
cg24556382	GALNT7	-0.46(-0.56,-0.36)	5.89E-18	-0.36(-0.45,-0.26)	8.14E-13
cg09935388	GFI1	-1.05(-1.22,-0.89)	1.92E-35	-0.89(-1.03,-0.75)	9.03E-32
cg12876356	GFI1	-0.60(-0.70,-0.50)	7.84E-31	-1.15(-1.37,-0.92)	1.39E-21
cg18146737	GFI1	-0.75(-0.88,-0.61)	4.26E-27	-0.66(-0.85,-0.47)	1.41E-11
cg18316974	GFI1	-0.60(-0.71,-0.48)	2.34E-25	-0.56(-0.71,-0.41)	1.46E-12
cg09662411	GFI1	-0.28(-0.33,-0.22)	5.15E-22	-0.56(-0.70,-0.42)	3.09E-14
cg06338710	GFI1	-0.50(-0.66,-0.34)	5.91E-10	-0.89(-1.16,-0.62)	2.16E-10
cg10399789	GFI1	-0.30(-0.40,-0.20)	2.76E-09	-0.33(-0.47,-0.19)	3.09E-06
cg14179389	GFI1	-0.23(-0.31,-0.15)	3.18E-08	-0.37(-0.49,-0.25)	3.01E-09
cg24741609	GLIS1	-0.20(-0.26,-0.14)	7.21E-12	-0.15(-0.22,-0.08)	1.89E-05
cg19717773	GNA12	-0.33(-0.41,-0.25)	3.99E-16	-0.25(-0.33,-0.17)	6.77E-09
cg18446336	GNA12	-0.42(-0.53,-0.31)	1.41E-13	-0.36(-0.47,-0.25)	1.04E-09
cg09658497	GNA12	-0.37(-0.48,-0.26)	5.63E-11	-0.40(-0.53,-0.27)	5.01E-09
cg25189904	GNG12	-0.65(-0.74,-0.56)	1.37E-46	-0.50(-0.59,-0.40)	3.91E-21
cg13399816	GNG12	-0.20(-0.25,-0.15)	5.64E-14	-0.14(-0.19,-0.09)	4.56E-08
cg13184736	GNG12	-0.36(-0.46,-0.27)	2.12E-13	-0.36(-0.48,-0.24)	9.32E-09
cg13185177	GP5	0.38(0.30,0.46)	4.93E-20	0.18(0.10,0.26)	1.58E-05
cg19859270	GPR15	-0.39(-0.46,-0.33)	1.17E-30	-0.43(-0.51,-0.36)	8.19E-27
cg19254163	GPR44	-0.25(-0.30,-0.20)	3.70E-25	-0.22(-0.27,-0.17)	3.41E-17
cg19827923	GPR55	-0.20(-0.25,-0.16)	3.11E-18	-0.14(-0.20,-0.09)	2.20E-06
cg10814005	GPR68	-0.18(-0.22,-0.13)	2.68E-14	-0.08(-0.12,-0.04)	6.01E-05
cg05875421	GPR68	-0.09(-0.13,-0.06)	3.82E-08	-0.09(-0.13,-0.05)	6.29E-06
cg18642234	GPX1	-0.16(-0.18,-0.13)	1.33E-28	-0.10(-0.14,-0.07)	6.56E-11
cg16255816	HAP1	-0.11(-0.13,-0.08)	2.12E-13	-0.12(-0.17,-0.07)	5.99E-07
cg03373393	HAP1	-0.11(-0.14,-0.08)	2.94E-12	-0.07(-0.09,-0.04)	3.19E-07
cg12729894	HCCA2;CTSD	-0.21(-0.27,-0.14)	1.29E-09	-0.19(-0.27,-0.11)	7.84E-06

cg24049493	HIVEP3	0.39(0.31,0.48)	4.68E-20	0.41(0.32,0.51)	3.59E-16
cg15542713	HIVEP3	0.51(0.40,0.63)	2.47E-17	0.63(0.49,0.77)	1.27E-17
cg16145216	HIVEP3	0.22(0.17,0.27)	3.42E-16	0.25(0.19,0.32)	3.54E-15
cg02583484	HNRNPA1	-0.25(-0.30,-0.19)	2.45E-16	-0.25(-0.32,-0.18)	6.56E-13
cg08472795	HS6ST1	-0.20(-0.25,-0.14)	2.83E-13	-0.16(-0.21,-0.11)	4.72E-09
cg03222009	HS6ST1	-0.15(-0.19,-0.10)	2.15E-09	-0.11(-0.17,-0.06)	3.94E-05
cg21121843	HTT	-0.24(-0.30,-0.17)	2.15E-12	-0.18(-0.25,-0.12)	8.33E-08
cg08763102	HTT	-0.13(-0.18,-0.09)	1.49E-09	-0.13(-0.18,-0.09)	9.22E-09
cg02968508	IL17RE	-0.22(-0.30,-0.15)	4.86E-09	-0.24(-0.33,-0.14)	6.47E-07
cg25809905	ITGA2B	0.31(0.22,0.40)	2.98E-11	0.23(0.14,0.33)	1.72E-06
cg09099830	ITGAL	-0.15(-0.18,-0.12)	8.51E-21	-0.12(-0.15,-0.08)	2.05E-09
cg06235438	ITGAL	-0.26(-0.33,-0.19)	2.60E-13	-0.19(-0.27,-0.11)	1.90E-06
cg16519923	ITGAL	-0.23(-0.29,-0.16)	1.92E-11	-0.17(-0.24,-0.10)	8.55E-06
cg06972908	ITGAL	-0.17(-0.23,-0.12)	4.33E-09	-0.13(-0.19,-0.08)	3.59E-06
cg05284742	ITPK1	-0.33(-0.37,-0.28)	3.00E-43	-0.26(-0.31,-0.22)	5.45E-26
cg17416793	KCNQ1	-0.14(-0.20,-0.09)	9.58E-08	-0.12(-0.18,-0.07)	2.43E-06
cg26963277	KCNQ1OT1;KCNQ1	-0.42(-0.50,-0.35)	2.18E-28	-0.38(-0.48,-0.28)	1.58E-13
cg16556677	KCNQ1OT1;KCNQ1	-0.37(-0.44,-0.29)	5.45E-23	-0.33(-0.42,-0.25)	1.23E-13
cg01744331	KCNQ1OT1;KCNQ1	-0.29(-0.36,-0.22)	1.10E-16	-0.29(-0.37,-0.21)	4.02E-12
cg07123182	KCNQ1OT1;KCNQ1	-0.27(-0.33,-0.20)	5.79E-16	-0.29(-0.36,-0.22)	6.85E-14
cg13708645	KDM2B	-0.15(-0.19,-0.10)	7.69E-11	-0.15(-0.21,-0.08)	3.70E-06
cg26995224	KDM2B	-0.24(-0.32,-0.17)	4.34E-10	-0.27(-0.37,-0.18)	6.11E-08
cg02451831	KIAA0087	-0.33(-0.40,-0.26)	6.76E-20	-0.27(-0.34,-0.21)	6.59E-16
cg22313519	KIAA1683	-0.15(-0.19,-0.10)	4.35E-10	-0.12(-0.17,-0.08)	1.21E-07
cg10255761	KLHDC8B	-0.17(-0.22,-0.12)	4.78E-10	-0.18(-0.25,-0.11)	2.39E-07
cg25420507	LGALS7	-0.12(-0.16,-0.08)	1.26E-08	-0.08(-0.12,-0.05)	7.15E-06
cg22649124	LGALS7B	-0.12(-0.15,-0.08)	8.90E-10	-0.10(-0.13,-0.06)	1.93E-07
cg00835193	LINGO3	-0.82(-1.03,-0.62)	6.13E-15	-0.91(-1.18,-0.64)	1.28E-10
cg21869609	LINGO3	-0.31(-0.41,-0.22)	5.54E-11	-0.26(-0.39,-0.14)	2.83E-05
cg01294327	LINGO3	-0.29(-0.37,-0.20)	6.94E-11	-0.21(-0.31,-0.11)	3.02E-05
cg07251887	RECQL5	-0.35(-0.40,-0.29)	6.68E-31	-0.24(-0.30,-0.18)	1.39E-15
cg03489965	LOC390594	0.24(0.18,0.30)	7.28E-14	0.22(0.13,0.31)	4.93E-06
cg10619342	LOC390594	0.30(0.22,0.38)	1.59E-13	0.26(0.14,0.39)	4.85E-05
cg21611682	LRP5	-0.33(-0.37,-0.29)	2.52E-52	-0.33(-0.38,-0.27)	1.09E-31
cg14624207	LRP5	-0.33(-0.38,-0.27)	6.19E-34	-0.24(-0.30,-0.19)	5.25E-16
cg10420527	LRP5	-0.21(-0.25,-0.17)	1.17E-25	-0.16(-0.20,-0.12)	9.94E-15
cg09578155	LRP5	-0.15(-0.19,-0.11)	3.93E-14	-0.13(-0.18,-0.09)	4.20E-09
cg07202214	LRRC32	-0.12(-0.15,-0.09)	9.94E-17	-0.06(-0.09,-0.04)	1.62E-05
cg13985437	LRRC32	-0.11(-0.14,-0.08)	7.41E-15	-0.07(-0.10,-0.04)	1.49E-05
cg10788371	LRRC32	-0.16(-0.20,-0.12)	1.45E-15	-0.15(-0.20,-0.10)	2.50E-08
cg13633560	LRRC32	-0.09(-0.12,-0.06)	5.49E-10	-0.08(-0.11,-0.04)	3.92E-05
cg09837977	LRRN3;IMMP2L	-0.23(-0.31,-0.15)	1.27E-08	-0.23(-0.32,-0.15)	2.48E-07
cg12423733	MAS1L	0.23(0.17,0.30)	1.50E-12	0.14(0.08,0.20)	4.06E-06
cg01435643	MCF2L	0.18(0.11,0.25)	9.57E-08	0.19(0.11,0.26)	1.63E-06
cg04468081	MCF2L2;B3GNT5	-0.14(-0.19,-0.09)	1.37E-08	-0.11(-0.17,-0.06)	6.29E-05
cg19918734	ME3	-0.12(-0.15,-0.09)	3.35E-12	-0.06(-0.09,-0.03)	4.10E-05

cg02556393	MECOM	-0.11(-0.15,-0.07)	8.77E-09	-0.09(-0.13,-0.04)	5.42E-05
cg20438687	MINK1	-0.21(-0.28,-0.13)	1.75E-08	-0.18(-0.25,-0.12)	5.75E-08
cg00687674	MIR548H4;TMEM84	0.18(0.12,0.25)	5.78E-08	0.20(0.11,0.29)	1.49E-05
cg00871610	MIR802	-0.23(-0.31,-0.15)	9.27E-09	-0.23(-0.30,-0.16)	1.51E-09
cg07381806	MOBKL2A	-0.32(-0.39,-0.25)	3.00E-17	-0.23(-0.31,-0.16)	1.31E-09
cg15187398	MOBKL2A	-0.21(-0.26,-0.16)	7.01E-16	-0.17(-0.22,-0.12)	1.75E-11
cg11621113	MORG1;MAN2B1	-0.11(-0.13,-0.08)	3.43E-17	-0.05(-0.08,-0.03)	6.52E-05
cg23842572	MPRIP	0.15(0.11,0.19)	2.86E-14	0.13(0.08,0.18)	4.93E-07
cg24838345	MTSS1	-0.34(-0.44,-0.24)	2.59E-11	-0.36(-0.47,-0.25)	2.41E-10
cg22132788	MYO1G	0.87(0.74,1.00)	2.39E-37	0.73(0.59,0.88)	9.47E-22
cg07826859	MYO1G	-0.28(-0.33,-0.24)	1.57E-35	-0.22(-0.27,-0.17)	1.08E-16
cg12803068	MYO1G	1.16(0.97,1.36)	2.80E-30	1.06(0.84,1.28)	2.84E-19
cg04180046	MYO1G	0.13(0.10,0.17)	3.19E-13	0.18(0.13,0.23)	2.45E-11
cg19089201	MYO1G	0.35(0.25,0.45)	2.29E-12	0.35(0.25,0.46)	8.28E-11
cg04039799	NAV2;NAV2;NAV2	-0.17(-0.22,-0.12)	5.55E-11	-0.14(-0.19,-0.09)	8.45E-08
cg03147185	NCAPH	-0.14(-0.18,-0.09)	4.02E-08	-0.10(-0.15,-0.05)	5.96E-05
cg02532700	NCF4;NCF4	-0.25(-0.30,-0.19)	5.06E-17	-0.17(-0.23,-0.11)	1.43E-07
cg06595162	NCRNA00114	-0.26(-0.32,-0.20)	1.07E-15	-0.20(-0.26,-0.14)	5.32E-10
cg26271591	NFE2L2	-0.43(-0.53,-0.33)	6.45E-17	-0.41(-0.52,-0.30)	5.17E-12
cg14120703	NOTCH1	-0.08(-0.11,-0.06)	4.47E-11	-0.07(-0.10,-0.04)	4.62E-07
cg05396397	NPPA	0.10(0.06,0.13)	2.05E-08	0.09(0.05,0.13)	3.34E-05
cg03340878	OR2B6	-0.31(-0.42,-0.20)	5.72E-08	-0.35(-0.47,-0.24)	2.60E-09
cg23126342	PCDH9	0.34(0.23,0.44)	1.17E-10	0.36(0.23,0.49)	5.49E-08
cg25491122	PCDH9	0.31(0.20,0.42)	3.45E-08	0.31(0.17,0.45)	2.01E-05
cg13039251	PDZD2	0.40(0.29,0.50)	1.48E-13	0.33(0.20,0.47)	6.67E-07
cg16151960	PHF15	-0.12(-0.16,-0.08)	4.76E-10	-0.09(-0.13,-0.05)	3.69E-06
cg21280392	PHOSPHO1	0.10(0.06,0.14)	8.67E-08	0.10(0.06,0.13)	8.12E-08
cg05460226	PIK3R5	-0.30(-0.37,-0.23)	2.70E-16	-0.21(-0.29,-0.13)	4.27E-07
cg16503724	PLCL2	0.24(0.18,0.30)	1.28E-15	0.14(0.09,0.20)	7.50E-07
cg05673882	POLK	-0.31(-0.37,-0.24)	8.30E-19	-0.25(-0.31,-0.18)	6.26E-12
cg26669717	PRKAR1B	-0.12(-0.15,-0.08)	5.11E-10	-0.09(-0.13,-0.05)	6.26E-06
cg11660018	PRSS23	-0.30(-0.34,-0.27)	4.80E-51	-0.27(-0.31,-0.22)	1.01E-24
cg23771366	PRSS23	-0.24(-0.27,-0.20)	2.97E-34	-0.21(-0.26,-0.16)	3.02E-15
cg23351584	PRSS23	-0.17(-0.20,-0.13)	4.53E-19	-0.09(-0.12,-0.06)	3.98E-08
cg12075928	PTK2	-0.39(-0.47,-0.31)	2.20E-20	-0.34(-0.43,-0.25)	2.46E-13
cg05668853	RAB34	-0.11(-0.15,-0.08)	2.52E-10	-0.11(-0.15,-0.06)	3.67E-05
cg19572487	RARA	-0.54(-0.62,-0.46)	2.53E-40	-0.48(-0.57,-0.38)	2.54E-21
cg11094248	RARA	-0.08(-0.11,-0.05)	2.45E-09	-0.08(-0.11,-0.06)	2.65E-09
cg10062919	RARA	-0.11(-0.13,-0.08)	9.12E-18	-0.07(-0.10,-0.04)	7.59E-06
cg05824218	RARA	0.09(0.07,0.12)	8.77E-12	0.09(0.05,0.12)	9.84E-07
cg04956244	RARA	0.10(0.07,0.14)	2.80E-10	0.09(0.06,0.13)	2.83E-08
cg16704246	RBM20	-0.19(-0.25,-0.13)	1.03E-09	-0.14(-0.21,-0.07)	3.79E-05
cg16969872	RBM26	-0.29(-0.38,-0.21)	1.61E-11	-0.21(-0.29,-0.13)	4.95E-07
cg23913963	RCC2	-0.28(-0.37,-0.18)	2.52E-08	-0.25(-0.36,-0.13)	5.02E-05
cg00214171	RECQL5;LOC100130933	-0.10(-0.13,-0.07)	8.07E-09	-0.12(-0.16,-0.08)	9.21E-10
cg11701312	RPS5	-0.12(-0.15,-0.09)	1.96E-13	-0.09(-0.13,-0.05)	1.01E-05

cg10951873	RUNX3	-0.12(-0.16,-0.08)	1.18E-08	-0.08(-0.11,-0.04)	9.10E-06
cg00310412	SEMA7A	-0.18(-0.22,-0.14)	3.79E-21	-0.18(-0.22,-0.13)	1.44E-14
cg18335991	SEMA7A	-0.27(-0.34,-0.20)	1.59E-14	-0.20(-0.28,-0.13)	6.12E-08
cg26908328	SERINC5	-0.14(-0.18,-0.10)	6.08E-10	-0.11(-0.16,-0.07)	7.79E-07
cg10531355	SERINC5	-0.22(-0.30,-0.15)	1.38E-09	-0.16(-0.23,-0.09)	7.49E-06
cg26856289	SFRS13A	-0.15(-0.18,-0.11)	7.89E-16	-0.10(-0.13,-0.06)	2.62E-08
cg14270346	SHB	-0.22(-0.30,-0.15)	2.38E-09	-0.20(-0.29,-0.11)	1.18E-05
cg14712058	SIN3B	-0.13(-0.17,-0.09)	2.41E-09	-0.13(-0.18,-0.08)	1.05E-06
cg01979157	SKI	-0.24(-0.31,-0.17)	3.85E-12	-0.19(-0.27,-0.10)	3.94E-05
cg08884752	SKI	-0.19(-0.25,-0.13)	6.17E-10	-0.18(-0.24,-0.13)	1.44E-09
cg09469355	SKI	-0.11(-0.15,-0.07)	3.41E-08	-0.14(-0.18,-0.09)	5.70E-09
cg07626482	SLC1A5	-0.08(-0.10,-0.05)	1.54E-09	-0.07(-0.10,-0.04)	3.03E-05
cg16547579	SLC23A2	-0.14(-0.18,-0.09)	2.13E-09	-0.09(-0.13,-0.04)	4.96E-05
cg25212453	SLC43A2	-0.18(-0.23,-0.13)	2.23E-11	-0.16(-0.23,-0.10)	2.77E-06
cg09197783	SLC43A3	-0.22(-0.29,-0.16)	5.72E-11	-0.18(-0.23,-0.13)	2.57E-11
cg05438378	SMAD3	-0.17(-0.22,-0.11)	9.75E-09	-0.13(-0.19,-0.08)	9.75E-06
cg00604410	SMAD3	-0.15(-0.20,-0.10)	8.79E-09	-0.13(-0.18,-0.08)	3.32E-06
cg01763916	SMAP2	-0.28(-0.33,-0.22)	1.06E-21	-0.14(-0.19,-0.09)	2.25E-07
cg13916835	SMG6	-0.29(-0.40,-0.19)	2.58E-08	-0.33(-0.43,-0.23)	2.01E-10
cg26718213	SNED1	0.19(0.12,0.25)	1.25E-08	0.26(0.18,0.33)	7.06E-11
cg26707709	SNED1	0.23(0.15,0.30)	3.18E-08	0.27(0.17,0.37)	1.34E-07
cg27312979	SORBS1	0.27(0.21,0.32)	7.26E-19	0.18(0.12,0.24)	1.00E-08
cg25722983	STK40	-0.16(-0.21,-0.12)	5.03E-12	-0.10(-0.15,-0.05)	2.74E-05
cg07465627	STXBP4	-0.15(-0.20,-0.10)	3.45E-09	-0.13(-0.18,-0.09)	1.05E-08
cg26790897	SUMF2	0.23(0.15,0.31)	2.98E-09	0.16(0.08,0.24)	6.59E-05
cg26701785	SYNJ2	0.29(0.21,0.38)	1.87E-11	0.23(0.12,0.33)	1.89E-05
cg11071448	SYT2	-0.25(-0.32,-0.18)	5.44E-13	-0.23(-0.29,-0.16)	4.99E-11
cg22966895	TAPBP;RGL2	-0.08(-0.11,-0.05)	2.59E-08	-0.08(-0.11,-0.05)	2.10E-06
cg06819357	TECPR2	0.25(0.17,0.32)	3.14E-11	0.15(0.09,0.21)	5.63E-07
cg05886626	THBS1	-0.16(-0.21,-0.11)	8.95E-11	-0.07(-0.10,-0.03)	6.85E-05
cg22870429	TIGIT	-0.15(-0.21,-0.10)	9.13E-08	-0.13(-0.18,-0.08)	5.22E-07
cg07180646	TMEM51	-0.25(-0.31,-0.19)	3.64E-16	-0.17(-0.24,-0.09)	1.54E-05
cg21913886	TMEM51	-0.37(-0.47,-0.27)	1.58E-13	-0.35(-0.46,-0.24)	1.68E-09
cg26729380	TNF	-0.18(-0.23,-0.13)	3.80E-12	-0.10(-0.14,-0.05)	6.65E-05
cg08553327	TNF	-0.15(-0.19,-0.10)	2.11E-10	-0.08(-0.12,-0.04)	6.85E-05
cg21222743	TNF	-0.16(-0.21,-0.11)	3.53E-10	-0.11(-0.15,-0.06)	7.19E-06
cg04425624	TNF	-0.12(-0.16,-0.08)	6.19E-08	-0.08(-0.11,-0.04)	3.14E-05
cg07094298	TNIP2	-0.21(-0.27,-0.14)	3.06E-09	-0.17(-0.24,-0.10)	2.83E-06
cg09022230	TNRC18	-0.39(-0.47,-0.30)	1.40E-17	-0.38(-0.46,-0.29)	6.34E-17
cg15022400	TRIM69	-0.17(-0.22,-0.13)	5.12E-15	-0.08(-0.11,-0.04)	2.57E-05
cg22851200	TRIP6	-0.16(-0.21,-0.11)	7.49E-10	-0.18(-0.25,-0.10)	3.95E-06
cg13525276	TSHR	0.28(0.21,0.36)	1.79E-12	0.25(0.18,0.33)	5.30E-10
cg10825315	TSHR	0.22(0.16,0.29)	3.33E-12	0.17(0.10,0.24)	4.99E-07
cg18625627	TSHR	0.22(0.15,0.29)	6.28E-10	0.21(0.14,0.28)	7.93E-10
cg20886049	TSKU	-0.22(-0.29,-0.14)	5.42E-08	-0.22(-0.31,-0.13)	4.00E-06
cg00295485	UXS1	-0.23(-0.31,-0.15)	1.25E-08	-0.21(-0.28,-0.13)	2.33E-07

cg17619755	VARS	0.22(0.17,0.28)	5.48E-15	0.16(0.10,0.23)	3.19E-07
cg10807309	VARS	0.17(0.11,0.23)	1.30E-08	0.18(0.11,0.26)	2.49E-06
cg05302489	VARS	0.24(0.15,0.33)	8.42E-08	0.21(0.11,0.30)	1.85E-05
cg19421584	WDR60	0.10(0.06,0.13)	8.35E-08	0.09(0.05,0.14)	3.32E-05
cg12276019	XKR6	-0.17(-0.22,-0.12)	1.39E-11	-0.14(-0.18,-0.09)	1.65E-07
cg16794579	XYLT1	-0.21(-0.27,-0.15)	1.07E-12	-0.16(-0.21,-0.10)	1.64E-08
cg04359840	XYLT1	-0.25(-0.32,-0.18)	7.80E-12	-0.18(-0.26,-0.11)	2.47E-06
cg06321596	XYLT1	-0.24(-0.32,-0.16)	4.52E-09	-0.25(-0.32,-0.17)	2.12E-10
cg26361535	ZC3H3	-0.42(-0.50,-0.33)	3.08E-22	-0.40(-0.50,-0.29)	3.76E-13
cg27262054	ZFHX3	-0.14(-0.18,-0.10)	3.59E-11	-0.10(-0.15,-0.06)	1.18E-05
cg01062937	ZFPM1	0.05(0.03,0.07)	1.35E-08	0.08(0.05,0.11)	7.10E-07
cg14977938	ZFYVE21	0.22(0.17,0.27)	1.81E-17	0.14(0.07,0.20)	3.02E-05
cg26242531	ZFYVE21	0.21(0.16,0.27)	6.17E-14	0.20(0.14,0.27)	6.43E-09
cg02743070	ZMIZ1	-0.20(-0.24,-0.15)	4.38E-17	-0.15(-0.20,-0.11)	1.51E-09
cg17823346	ZMIZ1	-0.19(-0.25,-0.13)	6.12E-10	-0.19(-0.26,-0.12)	3.50E-07
cg02145310	ZMIZ1	-0.10(-0.13,-0.07)	9.26E-10	-0.08(-0.11,-0.04)	2.72E-05
cg12303084	ZMYND8	-0.17(-0.21,-0.13)	1.77E-14	-0.10(-0.15,-0.06)	1.25E-06
cg11824827	ZNF668	0.13(0.09,0.17)	9.38E-11	0.18(0.12,0.24)	9.03E-09
cg21566642		-0.97(-1.03,-0.91)	4.07E-160	-0.79(-0.88,-0.70)	1.37E-51
cg01940273		-0.53(-0.57,-0.49)	1.18E-136	-0.45(-0.50,-0.40)	5.99E-55
cg05951221		-0.48(-0.51,-0.44)	2.77E-124	-0.43(-0.47,-0.38)	3.20E-52
cg06126421		-1.28(-1.38,-1.17)	3.45E-107	-1.01(-1.13,-0.89)	4.79E-47
cg03329539		-0.29(-0.32,-0.26)	5.67E-70	-0.25(-0.28,-0.21)	5.57E-34
cg14753356		-0.51(-0.56,-0.45)	1.85E-65	-0.49(-0.55,-0.42)	3.07E-37
cg24859433		-0.74(-0.82,-0.65)	6.12E-64	-0.63(-0.73,-0.52)	4.12E-29
cg15342087		-0.70(-0.79,-0.62)	2.01E-57	-0.74(-0.84,-0.63)	1.63E-37
cg27241845		-0.53(-0.60,-0.46)	4.36E-46	-0.47(-0.56,-0.39)	3.04E-24
cg06644428		-0.71(-0.81,-0.61)	6.78E-44	-0.67(-0.80,-0.55)	5.88E-23
cg23079012		-0.47(-0.54,-0.40)	1.04E-36	-0.44(-0.56,-0.33)	2.54E-14
cg04885881		-0.62(-0.71,-0.52)	1.12E-36	-0.47(-0.57,-0.36)	2.37E-17
cg00073090		-0.10(-0.12,-0.08)	9.01E-30	-0.09(-0.11,-0.06)	1.94E-12
cg27537125		-0.22(-0.25,-0.18)	1.89E-29	-0.15(-0.18,-0.12)	2.35E-20
cg13193840		-0.33(-0.39,-0.27)	5.79E-28	-0.25(-0.33,-0.17)	6.82E-10
cg03274391		0.98(0.81,1.16)	1.28E-26	0.73(0.53,0.92)	8.43E-13
cg23480021		1.03(0.83,1.23)	7.35E-24	0.96(0.71,1.22)	1.08E-12
cg01208318		-0.59(-0.70,-0.47)	5.70E-22	-0.46(-0.59,-0.32)	1.27E-10
cg12547807		-0.22(-0.27,-0.18)	5.76E-21	-0.15(-0.21,-0.10)	3.74E-09
cg08035323		0.41(0.33,0.50)	6.42E-21	0.24(0.14,0.33)	1.31E-06
cg13038618		-0.39(-0.47,-0.31)	1.26E-20	-0.33(-0.42,-0.24)	2.46E-12
cg01765406		-0.23(-0.28,-0.19)	2.06E-20	-0.22(-0.27,-0.16)	1.96E-14
cg24540678		-0.14(-0.17,-0.11)	4.46E-20	-0.10(-0.13,-0.07)	1.78E-11
cg01127300		-0.35(-0.43,-0.28)	7.09E-20	-0.28(-0.39,-0.18)	7.21E-08
cg12147622		-0.27(-0.33,-0.21)	2.37E-19	-0.17(-0.24,-0.11)	1.57E-07
cg05194346		0.41(0.32,0.50)	4.16E-19	0.23(0.15,0.32)	3.50E-07
cg12513616		-0.12(-0.14,-0.09)	9.11E-19	-0.10(-0.14,-0.07)	3.46E-11
cg01513913		-0.26(-0.32,-0.21)	1.32E-18	-0.23(-0.31,-0.15)	9.83E-09

cg21393163	-0.27(-0.32,-0.21)	3.09E-18	-0.15(-0.21,-0.09)	1.80E-06
cg03547355	-0.22(-0.26,-0.17)	2.42E-17	-0.17(-0.22,-0.12)	2.82E-11
cg07069636	-0.14(-0.17,-0.10)	2.46E-15	-0.10(-0.14,-0.06)	1.57E-06
cg15693572	0.55(0.41,0.69)	7.07E-15	0.48(0.32,0.64)	3.66E-09
cg01692968	-0.18(-0.23,-0.14)	7.88E-15	-0.13(-0.17,-0.09)	1.09E-10
cg16611234	-0.17(-0.21,-0.13)	1.02E-14	-0.09(-0.13,-0.06)	8.21E-07
cg20059928	-0.26(-0.33,-0.20)	2.78E-14	-0.32(-0.41,-0.23)	1.49E-11
cg23594345	-0.45(-0.57,-0.33)	8.88E-14	-0.47(-0.61,-0.33)	1.13E-10
cg08266095	-0.35(-0.44,-0.25)	5.76E-13	-0.24(-0.32,-0.15)	6.00E-08
cg25305703	-0.36(-0.46,-0.26)	1.74E-12	-0.41(-0.51,-0.30)	9.63E-14
cg11827514	-0.49(-0.63,-0.36)	2.75E-12	-0.42(-0.57,-0.27)	5.46E-08
cg02985540	-0.14(-0.18,-0.10)	3.56E-12	-0.12(-0.16,-0.07)	6.46E-08
cg02964434	-0.21(-0.27,-0.15)	6.52E-12	-0.12(-0.17,-0.07)	2.14E-06
cg03399898	-0.18(-0.23,-0.13)	8.26E-12	-0.15(-0.21,-0.08)	4.58E-06
cg19589396	-0.26(-0.33,-0.19)	1.02E-11	-0.16(-0.23,-0.09)	1.45E-05
cg22539182	0.10(0.07,0.13)	1.97E-11	0.09(0.06,0.13)	1.25E-07
cg19825437	-0.22(-0.28,-0.15)	3.03E-11	-0.23(-0.32,-0.15)	6.68E-08
cg13518625	-0.19(-0.25,-0.14)	4.52E-11	-0.22(-0.28,-0.16)	2.08E-12
cg19719391	0.21(0.15,0.27)	5.03E-11	0.13(0.07,0.19)	2.32E-05
cg05339037	-0.13(-0.17,-0.09)	5.46E-11	-0.12(-0.18,-0.07)	6.33E-06
cg21733098	-0.46(-0.59,-0.32)	6.86E-11	-0.38(-0.53,-0.24)	2.88E-07
cg23233742	-0.19(-0.25,-0.13)	8.14E-11	-0.17(-0.23,-0.10)	4.58E-07
cg09974965	-0.20(-0.26,-0.14)	1.26E-10	-0.13(-0.19,-0.07)	2.19E-05
cg21140898	-0.20(-0.26,-0.14)	1.82E-10	-0.17(-0.23,-0.10)	4.39E-07
cg12873476	-0.13(-0.17,-0.09)	2.17E-10	-0.11(-0.16,-0.06)	1.65E-05
cg12593793	-0.12(-0.15,-0.08)	2.73E-10	-0.08(-0.12,-0.05)	7.85E-06
cg01882991	-0.10(-0.13,-0.07)	5.21E-10	-0.09(-0.12,-0.06)	3.90E-10
cg24448421	0.20(0.13,0.26)	1.39E-09	0.20(0.13,0.27)	7.12E-09
cg23681440	-0.22(-0.30,-0.15)	1.62E-09	-0.19(-0.26,-0.12)	1.41E-07
cg16201146	-0.26(-0.34,-0.17)	3.09E-09	-0.21(-0.29,-0.12)	1.51E-06
cg15410835	-0.35(-0.47,-0.23)	3.77E-09	-0.32(-0.46,-0.19)	2.98E-06
cg13787850	-0.23(-0.31,-0.15)	4.31E-09	-0.29(-0.38,-0.21)	2.58E-11
cg00541303	-0.22(-0.29,-0.15)	4.69E-09	-0.15(-0.21,-0.09)	1.62E-06
cg23635560	-0.13(-0.17,-0.09)	5.53E-09	-0.11(-0.16,-0.07)	9.81E-07
cg19427338	0.13(0.09,0.18)	6.30E-09	0.13(0.07,0.18)	2.20E-05
cg19372602	-0.22(-0.29,-0.14)	1.22E-08	-0.17(-0.25,-0.09)	3.58E-05
cg27449150	-0.11(-0.15,-0.07)	1.30E-08	-0.10(-0.14,-0.06)	3.22E-07
cg13074055	-0.38(-0.50,-0.25)	1.39E-08	-0.48(-0.63,-0.33)	5.33E-10
cg00980649	-0.19(-0.26,-0.12)	1.91E-08	-0.15(-0.21,-0.08)	3.54E-05
cg23090529	-0.20(-0.27,-0.13)	3.27E-08	-0.18(-0.24,-0.11)	1.13E-07
cg05500734	-0.16(-0.21,-0.10)	4.52E-08	-0.15(-0.22,-0.08)	1.63E-05
cg17907003	-0.19(-0.26,-0.12)	4.52E-08	-0.23(-0.29,-0.17)	1.42E-12
cg02787737	0.14(0.09,0.19)	5.22E-08	0.13(0.08,0.19)	3.11E-06
cg00116430	-0.09(-0.13,-0.06)	9.38E-08	-0.11(-0.15,-0.08)	5.26E-09

Table S3 Mediation analysis of smoking associated methylation with gene expression as mediator in F4 and F3

Mediation analysis is conducted using linear regression model adjusted for age, sex, BMI, alcohol consumption, metabolite ~ α'*CpG + γ*expression. α' is indicated as Smoking-methylation association; γ indicates gene expression- methylation association in the table.

S4 expression		Cpg site		Mediation effect	*P*	Smoking-methylation α'Association (95% CI)	*P*	Expression-methylation γ Association (95% CI)	*P*
ILMN_1710326	CLDND1	cg16519923	ITGAL	-0.12(-0.19,-0.06)	3.57E-05	-0.05(-0.26,0.16)	0.62	-0.44(-0.6,-0.28)	7.10E-08
ILMN_1773650	LRRN3	cg22635096	ADARB1	0.24(0.16,0.33)	6.86E-09	0.02(-0.15,0.19)	0.79	0.20(0.14,0.26)	4.82E-10
ILMN_1773650	LRRN3	cg10841124	AHRR	0.17(0.1,0.24)	2.64E-06	0.07(-0.08,0.23)	0.35	0.14(0.08,0.2)	1.43E-06
ILMN_1773650	LRRN3	cg03234777	AMICA1	-0.14(-0.19,-0.08)	5.49E-07	-0.13(-0.24,-0.01)	0.03	-0.11(-0.16,-0.07)	1.93E-07
ILMN_1773650	LRRN3	cg00893603	ATP8A2	-0.35(-0.48,-0.21)	2.24E-07	-0.10(-0.38,0.18)	0.48	-0.29(-0.39,-0.18)	5.97E-08
ILMN_1773650	LRRN3	cg01731783	C14orf43	-0.18(-0.24,-0.13)	2.44E-11	0.08(-0.02,0.19)	0.13	-0.15(-0.19,-0.11)	7.82E-14
ILMN_1773650	LRRN3	cg22851561	C14orf43	-0.21(-0.29,-0.13)	4.87E-08	-0.06(-0.22,0.1)	0.47	-0.17(-0.23,-0.12)	7.57E-09
ILMN_1773650	LRRN3	cg15159987	CPAMD8	-0.19(-0.25,-0.13)	3.18E-11	-0.02(-0.13,0.1)	0.76	-0.16(-0.2,-0.12)	1.22E-13
ILMN_1773650	**LRRN3**	**cg00501876**	**CSRNP1**	-0.21(-0.27,-0.16)	9.08E-14	-0.04(-0.15,0.06)	0.42	-0.18(-0.22,-0.14)	3.29E-18
ILMN_1773650	LRRN3	cg14099685	CUGBP1	-0.15(-0.21,-0.1)	6.62E-09	0.01(-0.1,0.12)	0.87	-0.13(-0.17,-0.09)	4.58E-10
ILMN_1773650	LRRN3	cg14179389	GFI1	0.32(0.22,0.42)	9.94E-11	-0.48(-0.68,-0.29)	1.93E-06	0.27(0.2,0.34)	7.73E-13
ILMN_1773650	LRRN3	cg24741609	GLIS1	-0.25(-0.33,-0.18)	2.71E-11	0.09(-0.06,0.25)	0.22	-0.21(-0.27,-0.16)	6.04E-14
ILMN_1773650	LRRN3	cg13708645	KDM2B	-0.10(-0.15,-0.05)	4.55E-05	-0.06(-0.17,0.05)	0.29	-0.08(-0.12,-0.04)	4.52E-05
ILMN_1773650	**LRRN3**	**cg09837977**	**LRRN3**	-0.48(-0.58,-0.37)	0	0.07(-0.1,0.24)	0.41	-0.40(-0.46,-0.34)	1.37E-32
ILMN_1773650	LRRN3	cg19918734	ME3	0.17(0.12,0.21)	1.67E-14	-0.26(-0.34,-0.18)	5.24E-10	0.14(0.11,0.17)	1.14E-19
ILMN_1773650	LRRN3	cg00687674	MIR548H4	-0.17(-0.24,-0.1)	2.16E-06	0.29(0.14,0.45)	2.01E-4	-0.14(-0.2,-0.08)	1.10E-06
ILMN_1773650	LRRN3	cg07381806	MOBKL2A	-0.27(-0.36,-0.19)	1.61E-10	-0.01(-0.18,0.16)	0.92	-0.23(-0.3,-0.17)	5.47E-13
ILMN_1773650	LRRN3	cg23842572	MPRIP	-0.09(-0.14,-0.05)	1.50E-05	0.32(0.23,0.42)	8.15E-11	-0.08(-0.11,-0.04)	1.20E-05
ILMN_1773650	LRRN3	cg04039799	NAV2	-0.19(-0.25,-0.12)	8.86E-09	0.07(-0.07,0.2)	0.33	-0.15(-0.2,-0.11)	6.96E-10
ILMN_1773650	**LRRN3**	**cg02532700**	**NCF4**	0.21(0.14,0.28)	2.43E-10	-0.53(-0.66,-0.4)	1.49E-14	0.17(0.13,0.22)	3.20E-12
ILMN_1773650	LRRN3	cg21280392	PHOSPHO1	0.14(0.1,0.18)	2.04E-11	-0.03(-0.11,0.05)	0.42	0.11(0.08,0.14)	5.79E-14

ILMN_1773650	LRRN3	cg05460226	PIK3R5	-0.24(-0.32,-0.16)	1.20E-08	-0.12(-0.29,0.06)	0.19	-0.20(-0.26,-0.14)	1.06E-09
ILMN_1773650	LRRN3	cg16503724	PLCL2	-0.19(-0.28,-0.11)	1.61E-06	0.47(0.3,0.65)	1.54E-07	-0.16(-0.22,-0.1)	7.68E-07
ILMN_1773650	**LRRN3**	**cg23771366**	**PRSS23**	-0.09(-0.13,-0.05)	3.40E-06	-0.18(-0.27,-0.1)	2.68E-05	-0.07(-0.11,-0.04)	1.96E-06
ILMN_1773650	LRRN3	cg11094248	RARA	0.06(0.03,0.1)	4.05E-05	-0.15(-0.22,-0.08)	3.72E-05	0.05(0.03,0.08)	3.94E-05
ILMN_1773650	LRRN3	cg16704246	RBM20	-0.28(-0.36,-0.2)	5.30E-12	0.02(-0.14,0.17)	0.85	-0.24(-0.29,-0.18)	5.73E-15
ILMN_1773650	LRRN3	cg16969872	RBM26	-0.19(-0.28,-0.1)	3.40E-05	-0.15(-0.35,0.05)	0.14	-0.16(-0.23,-0.08)	3.21E-05
ILMN_1773650	LRRN3	cg08884752	SKI	-0.38(-0.47,-0.28)	3.22E-15	0.16(-0.02,0.33)	0.08	-0.31(-0.38,-0.25)	3.69E-21
ILMN_1773650	LRRN3	cg16547579	SLC23A2	-0.17(-0.24,-0.1)	1.00E-06	-0.01(-0.16,0.14)	0.93	-0.14(-0.2,-0.09)	4.15E-07
ILMN_1773650	LRRN3	cg05438378	SMAD3	0.24(0.17,0.3)	1.34E-12	-0.39(-0.52,-0.27)	3.22E-09	0.20(0.15,0.25)	5.18E-16
ILMN_1773650	LRRN3	cg01763916	SMAP2	0.35(0.26,0.45)	2.33E-14	-0.64(-0.81,-0.47)	9.74E-13	0.30(0.23,0.36)	1.91E-19
ILMN_1773650	LRRN3	cg04425624	TNF	0.14(0.09,0.19)	2.89E-09	-0.30(-0.4,-0.21)	1.69E-09	0.12(0.08,0.15)	1.43E-10
ILMN_1773650	LRRN3	cg08553327	TNF	0.16(0.1,0.21)	1.97E-09	-0.34(-0.44,-0.23)	1.16E-09	0.13(0.09,0.17)	7.75E-11
ILMN_1773650	LRRN3	cg21222743	TNF	0.24(0.18,0.3)	2.49E-14	-0.37(-0.49,-0.25)	8.74E-10	0.20(0.16,0.24)	2.49E-19
ILMN_1773650	**LRRN3**	**cg26729380**	**TNF**	0.25(0.18,0.31)	6.14E-14	-0.51(-0.64,-0.39)	2.40E-15	0.21(0.16,0.25)	1.56E-18
ILMN_1773650	LRRN3	cg10825315	TSHR	-0.22(-0.32,-0.12)	7.68E-06	0.57(0.35,0.79)	3.35E-07	-0.18(-0.26,-0.11)	5.36E-06
ILMN_1773650	LRRN3	cg00295485	UXS1	-0.27(-0.36,-0.18)	1.55E-09	-0.03(-0.21,0.16)	0.78	-0.22(-0.29,-0.16)	4.98E-11
ILMN_1773650	LRRN3	cg00980649		-0.21(-0.29,-0.14)	2.43E-08	0.04(-0.12,0.2)	0.65	-0.18(-0.24,-0.12)	2.91E-09
ILMN_1773650	LRRN3	cg01208318		-0.30(-0.43,-0.17)	2.20E-06	-0.26(-0.53,0.02)	0.07	-0.25(-0.35,-0.15)	1.13E-06
ILMN_1773650	LRRN3	cg01513913		-0.16(-0.23,-0.09)	6.72E-06	-0.09(-0.24,0.06)	0.25	-0.13(-0.19,-0.08)	4.55E-06
ILMN_1773650	**LRRN3**	**cg00073090**		0.06(0.04,0.08)	3.56E-09	-0.17(-0.22,-0.13)	4.91E-16	0.05(0.03,0.06)	1.86E-10
ILMN_1773650	LRRN3	cg05339037		0.12(0.07,0.16)	1.80E-07	-0.22(-0.32,-0.13)	6.00E-06	0.10(0.06,0.13)	4.46E-08
ILMN_1773650	LRRN3	cg03547355		-0.19(-0.25,-0.13)	6.49E-11	-0.04(-0.16,0.07)	0.45	-0.16(-0.2,-0.11)	3.90E-13
ILMN_1773650	LRRN3	cg12547807		-0.19(-0.25,-0.12)	1.02E-08	-0.08(-0.21,0.06)	0.25	-0.15(-0.2,-0.11)	8.53E-10
ILMN_1773650	LRRN3	cg12593793		0.12(0.08,0.17)	7.51E-09	-0.24(-0.33,-0.15)	1.43E-07	0.10(0.07,0.14)	5.46E-10
ILMN_1773650	LRRN3	cg27449150		-0.10(-0.15,-0.05)	1.81E-05	0.01(-0.1,0.11)	0.91	-0.09(-0.13,-0.05)	1.52E-05
ILMN_1773650	LRRN3	cg27537125		-0.11(-0.16,-0.07)	7.54E-07	-0.09(-0.19,0)	0.06	-0.09(-0.13,-0.06)	2.91E-07
ILMN_1773650	**LRRN3**	**cg14753356**		0.18(0.11,0.24)	1.36E-07	-0.63(-0.77,-0.48)	2.60E-17	0.15(0.1,0.2)	3.06E-08

ILMN_1773650	LRRN3	cg24540678		-0.08(-0.11,-0.05)	9.87E-07	-0.07(-0.14,0)	0.04	-0.06(-0.09,-0.04)	4.11E-07
S2									
ILMN_1710326	CLDND1	cg16519923	ITGAL	-0.16(-0.52,0.19)'	0.37	0.08(-0.37,0.53)	0.38	0.01(-0.03,0.04)	0.73
ILMN_1773650	LRRN3	cg22635096	ADARB1	-0.07(-0.25,0.11)	0.22	0.3(0.02,0.57)	0.04	-0.11(-0.37,0.16)	0.43
ILMN_1773650	LRRN3	cg10841124	AHRR	0.01(-0.18,0.19)	0.48	0.12(-0.17,0.41)	0.42	0.01(-0.27,0.29)	0.95
ILMN_1773650	LRRN3	cg03234777	AMICA1	-0.01(-0.12,0.11)	0.46	-0.05(-0.22,0.13)	0.62	-0.01(-0.18,0.16)	0.92
ILMN_1773650	LRRN3	cg00893603	ATP8A2	0.09(-0.27,0.46)	0.31	-0.42(-0.98,0.15)	0.15	0.14(-0.4,0.69)	0.61
ILMN_1773650	LRRN3	cg01731783	C14orf43	0.03(-0.1,0.16)	0.31	-0.03(-0.23,0.18)	0.79	0.05(-0.15,0.25)	0.63
ILMN_1773650	LRRN3	cg22851561	C14orf43	0.08(-0.15,0.31)	0.25	-0.31(-0.66,0.05)	0.10	0.12(-0.23,0.46)	0.50
ILMN_1773650	LRRN3	cg15159987	CPAMD8	-0.02(-0.15,0.11)	0.38	-0.03(-0.24,0.17)	0.75	-0.03(-0.23,0.17)	0.75
ILMN_1773650	**LRRN3**	**cg00501876**	**CSRNP1**	0.02(-0.18,0.22)	0.41	-0.23(-0.54,0.08)	0.15	0.03(-0.27,0.33)	0.82
ILMN_1773650	LRRN3	cg14099685	CUGBP1	-0.01(-0.14,0.12)	0.44	-0.02(-0.22,0.18)	0.81	-0.01(-0.21,0.18)	0.88
ILMN_1773650	LRRN3	cg14179389	GFI1	0(-0.28,0.28)	0.49	-0.49(-0.93,-0.05)	0.03	0.01(-0.42,0.43)	0.98
ILMN_1773650	LRRN3	cg24741609	GLIS1	-0.01(-0.19,0.16)	0.45	-0.11(-0.38,0.17)	0.45	-0.02(-0.29,0.25)	0.89
ILMN_1773650	LRRN3	cg13708645	KDM2B	0.01(-0.15,0.16)	0.47	-0.16(-0.41,0.08)	0.20	0.01(-0.23,0.24)	0.95
ILMN_1773650	**LRRN3**	**cg09837977**	**LRRN3**	-0.09(-0.34,0.16)	0.24	-0.12(-0.51,0.26)	0.53	-0.13(-0.51,0.24)	0.48
ILMN_1773650	**LRRN3**	cg19918734	ME3	0.04(-0.06,0.13)	0.24	-0.08(-0.23,0.07)	0.28	0.05(-0.09,0.2)	0.47
ILMN_1773650	LRRN3	cg00687674	MIR548H4	-0.11(-0.3,0.08)	0.13	0.38(0.09,0.67)	0.01	-0.17(-0.45,0.11)	0.25
ILMN_1773650	**LRRN3**	cg07381806	MOBKL2A	0.02(-0.19,0.24)	0.41	-0.22(-0.57,0.13)	0.23	0.04(-0.35,0.43)	0.83
ILMN_1773650	LRRN3	cg23842572	MPRIP	-0.11(-0.27,0.04)	0.07	0.37(0.14,0.61)	0.00	-0.18(-0.4,0.05)	0.14
ILMN_1773650	LRRN3	cg04039799	NAV2	0.02(-0.14,0.18)	0.40	-0.12(-0.37,0.13)	0.34	0.03(-0.21,0.27)	0.79
ILMN_1773650	**LRRN3**	**cg02532700**	**NCF4**	0.08(-0.07,0.23)	0.15	-0.34(-0.56,-0.11)	0.01	0.12(-0.1,0.34)	0.29
ILMN_1773650	LRRN3	cg21280392	PHOSPHO1	0.1(0.01,0.19)	1.73E-02	0(-0.13,0.13)	0.97	0.15(0.02,0.27)	0.03
ILMN_1773650	LRRN3	cg05460226	PIK3R5	-0.01(-0.24,0.21)	0.45	-0.17(-0.52,0.19)	0.35	-0.02(-0.36,0.32)	0.91
ILMN_1773650	LRRN3	cg16503724	PLCL2	0.05(-0.13,0.23)	0.31	0.02(-0.26,0.3)	0.89	0.07(-0.2,0.34)	0.61
ILMN_1773650	**LRRN3**	**cg23771366**	**PRSS23**	0.06(-0.04,0.16)	0.12	-0.28(-0.44,-0.12)	1.20E-03	0.09(-0.06,0.25)	0.24

ILMN_1773650	LRRN3	cg11094248	RARA	0.04(-0.06,0.14)	0.21	-0.16(-0.32,-0.01)	0.04	0.06(-0.09,0.21)	0.42
ILMN_1773650	LRRN3	cg16704246	RBM20	-0.1(-0.28,0.08)	0.13	0.29(0.02,0.56)	0.04	-0.15(-0.42,0.11)	0.26
ILMN_1773650	LRRN3	cg16969872	RBM26	-0.03(-0.26,0.2)	0.39	-0.13(-0.49,0.23)	0.47	-0.05(-0.4,0.3)	0.79
ILMN_1773650	LRRN3	cg08884752	SKI	0.1(-0.1,0.3)	0.17	-0.19(-0.49,0.12)	0.24	0.15(-0.15,0.44)	0.33
ILMN_1773650	LRRN3	cg16547579	SLC23A2	-0.09(-0.24,0.07)	0.14	0.09(-0.15,0.33)	0.46	-0.13(-0.36,0.1)	0.27
ILMN_1773650	LRRN3	cg05438378	SMAD3	0.11(-0.04,0.26)	0.07	-0.28(-0.5,-0.05)	0.02	0.17(-0.05,0.39)	0.13
ILMN_1773650	LRRN3	cg01763916	SMAP2	0.09(-0.09,0.27)	0.16	-0.37(-0.65,-0.09)	0.01	0.14(-0.13,0.41)	0.32
ILMN_1773650	LRRN3	cg04425624	TNF	0.06(-0.03,0.15)	0.11	-0.17(-0.31,-0.03)	0.02	0.08(-0.05,0.22)	0.22
ILMN_1773650	LRRN3	cg08553327	TNF	0.04(-0.07,0.15)	0.23	-0.2(-0.37,-0.02)	0.03	0.06(-0.1,0.23)	0.47
ILMN_1773650	LRRN3	cg21222743	TNF	0(-0.12,0.12)	0.49	-0.09(-0.28,0.1)	0.36	0(-0.19,0.18)	0.98
ILMN_1773650	**LRRN3**	**cg26729380**	**TNF**	0.07(-0.05,0.2)	0.12	-0.23(-0.42,-0.04)	0.02	0.11(-0.07,0.3)	0.24
ILMN_1773650	LRRN3	cg10825315	TSHR	0.08(-0.19,0.34)	0.28	0.24(-0.16,0.65)	0.24	0.11(-0.28,0.51)	0.57
ILMN_1773650	LRRN3	cg00295485	UXS1	0.04(-0.2,0.28)	0.38	-0.02(-0.4,0.35)	0.90	0.06(-0.31,0.42)	0.77
ILMN_1773650	LRRN3	cg00980649		0.06(-0.13,0.24)	0.27	-0.23(-0.52,0.06)	0.13	0.09(-0.19,0.37)	0.55
ILMN_1773650	LRRN3	cg01208318		0.04(-0.3,0.39)	0.40	-0.45(-0.98,0.09)	0.11	0.07(-0.45,0.59)	0.80
ILMN_1773650	LRRN3	cg01513913		-0.02(-0.21,0.16)	0.40	-0.14(-0.43,0.14)	0.33	-0.04(-0.32,0.24)	0.80
ILMN_1773650	**LRRN3**	**cg00073090**		0.02(-0.04,0.08)	0.24	-0.14(-0.23,-0.05)	3.44E-03	0.03(-0.06,0.12)	0.48
ILMN_1773650	**LRRN3**	**cg05339037**		0.06(-0.07,0.19)	0.18	-0.27(-0.48,-0.06)	0.01	0.09(-0.11,0.29)	0.37
ILMN_1773650	LRRN3	cg03547355		0.04(-0.11,0.19)	0.29	-0.25(-0.48,-0.02)	0.04	0.06(-0.16,0.29)	0.59
ILMN_1773650	LRRN3	cg12547807		-0.01(-0.19,0.18)	0.47	-0.02(-0.31,0.27)	0.90	-0.01(-0.29,0.27)	0.95
ILMN_1773650	LRRN3	cg12593793		0.1(0,0.2)	2.36E-02	-0.25(-0.4,-0.1)	0.00	0.15(0.01,0.3)	0.04
ILMN_1773650	LRRN3	cg27449150		0.07(-0.07,0.2)	0.17	-0.19(-0.39,0.01)	0.07	0.1(-0.1,0.3)	0.33
ILMN_1773650	LRRN3	cg27537125		-0.01(-0.09,0.07)	0.42	-0.1(-0.23,0.03)	0.13	-0.01(-0.14,0.11)	0.84
ILMN_1773650	**LRRN3**	**cg14753356**		0.06(-0.1,0.23)	0.23	-0.5(-0.76,-0.24)	4.44E-04	0.1(-0.16,0.35)	0.46
ILMN_1773650	LRRN3	cg24540678		0(-0.07,0.07)	0.49	-0.09(-0.2,0.02)	0.11	0(-0.11,0.11)	0.98

Acknowledgement

First, I would like to thank Prof. Dr. Thomas Illig, CEO and scientific head of Hannover Unified Biobank (HUB), Hannover Medical School, former head of Research Unit of Molecular Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health who provided me this possibility to conduct the PhD study.

I would like to give my special thanks to my director supervisor Dr. Rui Wang-Sattler, head of the Group leader of Molecular Epidemiology – Metabolism Research Unit of Molecular Epidemiology, for her continuous support, great patience and invaluable suggestions throughout the projects. She made great efforts in coordinating the projects and also organized numerous discussions with experts in the field.

Thanks to Prof. Dr. Annette Peters for her great support in the metabolomics study of myocardial infarction, and Prof. Fabien Theis colleagues for his valuable insights and inspiring suggestions in my thesis committee.

The thesis would not be possible without the diligent work from MONICA/KORA studies (led by Prof. Dr. Dr. H.-Erich Wichmann and Prof. Dr. Annette Peters), who established the cohort, and the Genome Analysis Center (GAC) Metabolomic Platform (led by Prof. Dr. Jurek Adamsik), who made the metabolite profiling available. I would like to thank all the staffs in MONICA/KORA and GAC.

Many thanks to my current and former colleagues who gave me support during my PhD work and helped me get adapted to the life in Germany.

I am grateful for my family, my wife and my parents, for their continuous love, trust, and understanding which gave me strength to complete the PhD work.

Curriculum vitae

Tao Xu

Research Unit of Molecular Epidemiology
Helmholtz Zentrum München
D-85764 Neuherberg, Germany
E-mail: *tao.xu@helmholtz-muenchen.de*

Research interest

- Bioinformatics / Systems biology / Statistical modeling
- 'omics' data integration / Data visualization
- Metabolomics / Cardiovascular diseases
- Biomarker discovery

Education

2011–2014 **Promotion (Dr. rer. Nat.) in Epidemiology**, *Ludwig-Maximilians-Universität München*. Faculty of Medicine

2008–2011 **Master in Biomedical Engineering**, *Shanghai Jiao Tong University*. Department of Biostatistics and Bioinformatics, School of Life Science and Biotechnology

2004–2008 **Bachelor in Biomedical Engineering**, *Southeast University*. School of Biological Science and Medical Engineering

Research experience

2011– 2014 **PhD student**, *Research Unit Molecular Epidemiology, HelmholtzZentrum München*, Munich.

- Led metabolomics study of myocardial infarctions and hypertension in the group
- Led systems biology study of smoking in the group
- IT system administrator
- Participation in other projects, including biomarker discovery for diabetes, the study of nurse shift-work, etc.

2008–2011 **Working student**, *Bioinformatics Application Software R&D, Shanghai Center for Bioinformation Technology*, Shanghai.

- Software development, data analysis

Teaching

2011-2014 Teaching metabolomics and R in the master course "Genetic Epidemiology" in Ludwig-Maximilians-Universität München

Presentations and talks

2014	Workshop zur Genetischen Epidemiologie 2014, May 2014, Grainau, Germany
2013	21th Annual International Conference on Intelligent System for Molecular Biology (ISMB), July 2013, Berlin, Germany
2012	Workshop zur Genetischen Epidemiologie 2012, April 2012, Grainau, Germany
2010	Genome Informatics Workshop (GIW) 2010, December 2010, Hangzhou, China

Publications

1. Bernd Stratmann*, **Tao Xu***, Christa Meisinger*, … PLA1A2 platelet polymorphism predicts mortality in prediabetic subjects of the population based KORA S4-Cohort. *Cardiovascular Diabetology* 2014, 13:90
2. Marta Jaremek, Zhonghao Yu, ..., **Tao Xu**, ..., Thomas Illig, Tim D. Spector, and Rui Wang-Sattler. Alcohol-induced metabolomic differences in humans. *Translational Psychiatry*, 3(7):e276, July 2013.
3. **Tao Xu***, Xiao Chang*, Yun Li, and Kai Wang. Dynamic modular architecture of protein-protein interaction networks beyond the dichotomy of 'date' and 'party' hubs. *Scientific Reports*, 3, April 2013.
4. **Tao Xu**, Christina Holzapfel, Xiao Dong, ..., and Rui Wang-Sattler. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Medicine*, 11(1):60, March 2013. PMID: 23497222.
5. Rui Wang-Sattler, Zhonghao Yu, Christian Herder, ..., **Tao Xu**, ..., and Thomas Illig. Novel biomarkers for pre-diabetes identified by metabolomics. *Molecular systems biology*, 8:615, September 2012. PMID: 23010998.
6. Xiao Dong, Tingyan Zhong, **Tao Xu**, Yunting Xia, Biqing Li, Chao Li, Guohui Ding, and Yixue Li. Evaluating coverage of exons by HapMap SNPs. *Genomics*,September 2012.
7. Zhonghao Yu, Guangju Zhai, Paula Singmann, Ying He, **Tao Xu**, ..., Thomas Illig, and Rui Wang-Sattler. Human serum metabolic profiles are age dependent. *Aging cell*, July 2012. PMID: 22834969.
8. **Tao Xu***, Yao Yu*, Yongtao Yu, Pei Hao, and Xuan Li. Association of tissue lineage and gene expression: conservatively and differentially expressed genes define common and special functions of tissues. *BMC Bioinformatics*, 11(Suppl 11):S1, 2010.
9. **Tao Xu**, Jie Ping, Yao Yu, Fudong Yu, Yongtao Yu, Pei Hao, and Xuan Li. Revealing parasite influence in metabolic pathways in apicomplexa infected patients. *BMC Bioinformatics*, 11(Suppl 11):S13, 2010.
   *equal contribution