Conceptualisations of successful ageing and leads for lifestyle modification

Conceptualisations of successful ageing and leads for lifestyle modification

Sarah Mount

The research described in this thesis was performed at the Department of Respiratory Medicine, NUTRIM School of Nutrition and Translational Research in Metabolism at Maastricht University Medical Centre+, Maastricht, the Netherlands.

Cover design:

Layout by:

Printed by:

Published by:

ISBN:

Conceptualisations of successful ageing and leads for lifestyle modification

DISSERTATION

To obtain the degree of doctor at Maastricht University,

on the authority of the Rector Magnificus Prof.dr. Rianne M. Letschert

in accordance with the decision of the Board of Deans,

to be defended in public (date) (time).

by

Sarah Mount

born in North York Canada, February 2, 1987

Promotors

Prof. dr.ir. A.M.W.J. Schols

Prof.dr. M.P.A. Zeegers

Co-promotor

Dr. A. Wesselius

Assessment committee

Prof. dr. E.E. Blaak, Chair

Prof. dr. L. de Groot, Wageningen University

Prof. dr. A. Jansen

Prof. dr. A-H. Maitland van der Zee, Amsterdam UMC

Prof. dr. F-J. van Schooten

Table of contents

Chapter 1

General Introduction

To know where you are going, you need to know where you have been.

The relevance of ageing successfully is becoming increasingly important as life expectancy increases and birth rates fall. Successful ageing, at least according to the WHO is maintaining functional ability well into old age.[1] Between the years of 1965 and 2014 there were 20 million premature deaths, 10 million from metabolic, cardiovascular, neurovascular, and renovascular events and 6.9 million from respiratory diseases.[2] Disease-related premature death results in years of life lost directly, but it often also increases the number of years with reduced quality of life. Therefore, addressing risk factors for disability, frailty and dependency throughout the life course, becomes an important public health concern. In order to ensure positive actions towards this goal of disease prevention and increase the likelihood of healthy ageing, strides should be made to identify and mitigate risk factors. Risk factors for disease and dependency can be considered non-controllable (genetics, gender etc), distal (economic background, education, air pollution etc.), and intermediate (health habits: sedentary lifestyle, diet, alcohol consumption, smoking, healthcare access, living and working conditions). Unfortunately, it is not currently possible to target many non-controllable and distal risk factors that play important roles in disease development and healthy ageing. However, improvements in health assessments and interventions may allow us to limit the influence of these otherwise 'uncontrollable' risk factors.

Over the last century, we have quantified health risk using different risk scores as well as in the context of biological and chronological age. However, an operationalised definition of successful ageing has remained elusive. A recent citation network analysis revealed in excess of 5000 publications relating to successful ageing[3] published between 1902 and 2015. As of June 2019, an additional 2093 publications have been indexed by PubMed. The sheer volume of publications relating to this topic is indicative of the mounting problem society is facing due to falling birth rates and changes in the population pyramid; larger and larger proportions of adults becoming 65 years of age and older. Moreover, the older adult group is retired and often has increased medical needs. Effectively measuring and predicting who will age well, that is to say who will maintain their independence, functional abilities, emotional wellbeing, and who will not, is key in order to ensure the right person gets the right intervention at the right time. By developing an effective metric, we could not only predict who is likely to age successfully, but also to measure the effectiveness of interventions and better identify knowledge gaps.

Where we have been: Successful ageing

The aged are more prone to disease and have a more limited adaptive capacity when compared to adults. Approximately, 80 percent of the older adults have at least one age-related disorder, with 50 percent having at least two age related disorders.[4] For this reason ageing can be considered a risk factor for disease itself. However, ageing itself does not provide an explanation as to why age-related disorders do not develop in everyone nor why different individuals get different diseases.[4] Although over the last decades research has made strides to define successful ageing and to identify critical determinants of successful ageing, there is no consensus on the definition.[5]

The need to 'age successfully' is a modern phenomenon. In years prior, survival alone to an old age was an accomplishment, but today we demand more; we want longer and higher quality lives. However, the origins of the idea of successful ageing can be traced back to Rowe and Kahn's first model in 1987[6] which boiled down to a dichotomous model of individuals being classified as usual (non-diseased but high risk individuals) or successful (low risk individuals with high function) agers.[6] This model was revised in 1997 to define successful ageing as a combination of disease and disability avoidance, high cognitive and physical function as well as engagement with life.[7] However, concerns about the absolute dichotomous classification of individuals into successful and usual agers based on these criteria remained. Beyond the criticism of solely classifying individuals as successful or usual agers, these models focus on late adulthood and neglect to capture developmental processes and functional changes over time.[8] Newer models therefore increasingly incorporate a life course perspective which is a dynamic perspective that considers development, history and the importance of relationships over time. This incorporation offers the opportunity to understand successful ageing as a developmental process rather than a static process. In general, newer models tend to be multidimensional, however they also tend to be more appropriate for aged individuals and focus on frailty and mortality. Moreover, few models are comprehensive in including measures of successful ageing such as cognitive, physiological, psychological, physical as well as sensory capabilities and social wellbeing.

Research should explore the relationships between subjective and objective aspects of successful ageing and how they can be combined in order to make an operationalised measurement which is relevant to policy and practice.[9] We need to be mindful that successful

ageing can exist within the context of well managed chronic disease and multimorbidity, if high levels of functioning are maintained such that we do not only consider the "genetically fortunate".[10] Although no agreements have been made on the operationalisation of successful ageing, literature shows that there is a consistent association between midlife behaviours and late life outcomes.[10] Specifically leading a healthy life tends to promote longevity and good health. Therefore, physical activity, healthy diets, and smoking cessation efforts should be encouraged at all life stages in order to prevent disability and chronic disease.[10]

Chronic obstructive pulmonary disease (COPD): a disease model of accelerated ageing

As the prevalence of non-communicable diseases grows, there will be an increasing need for effective treatment and prevention. Poor lifestyle choices, such as tobacco smoking, excessive alcohol consumption, physical inactivity and diet are key contributors in the development of many chronic diseases and increasing evidence shows that these poor lifestyle choices, if addressed in mid-life, could substantially reduce the risk of the development of these diseases.

COPD is a lifestyle induced disease that shares common risk factors with other chronic diseases including diabetes and cardiovascular disease, although genetic predisposition also plays a role. The most important risk for COPD is tobacco smoking, but there is increasing evidence that air pollution and western style diet are also significant risk factors. The disease is characterised by persistent airflow limitation and respiratory symptoms[11] and unfortunately is currently the fourth leading cause of death world-wide.[12] Furthermore it is projected to become the third leading cause of (preventable and untimely) death by 2020.[13]

There is accumulating evidence that COPD exhibits many of the key hallmarks of ageing, and is therefore often used as a model of accelerated ageing.[14] Specifically, it is associated with telomere shortening, cellular senescence, activation of PI3 kinase-mTOR signalling, impaired autophagy, mitochondrial dysfunction, stem cell exhaustion, epigenetic changes, abnormal microRNA profiles, immune senescence, and low-grade systemic inflammation,[15] which are all common denominators of the human ageing process. Moreover, COPD has significant extra-pulmonary effects including weight loss, sarcopenia, nutritional abnormalities, skeletal muscle dysfunction, and is often associated with comorbidities including cardiovascular disease, diabetes, osteoporosis, depression, anxiety, cognitive impairment and cancer.[11] Therefore, COPD is an interesting model to study with respect to ageing as well as chronic disease development and progression.

There is no cure for COPD; [16] [17] as with many other non-communicable chronic diseases, making prevention a key goal.[18-20] The major aims of COPD management are to reduce risk factors, manage stable COPD and to manage associated comorbidities. In general, modern COPD therapies manage the individual[21] which emphasise pulmonary rehabilitation (PR), and physical activity[22] in addition to medications, lung volume reduction and smoking cessation.[23] However, maintaining an active lifestyle has historically been problematic in this patient population[24] [25] which has led interest in the development of behaviour-targeted interventions.[26] In patients and in individuals at risk of becoming patients of lifestyle related diseases, the barriers to risk reduction are the same, namely patient willingness, circumstance and knowledge.

Recent innovations, particularly in COPD management, include behavioural therapy, self-regulatory techniques, and motivational interviewing, but, continuous effort is still required to aid individuals in making advantageous health choices such as by improving their self-regulation. However, many of these types of healthcare provider driven methods are costly due to the need of counselling and patient monitoring. In addition, with growing number of individuals with disease and a declining workforce, other or additional approaches such as cognitive training may be better suited in inducing behaviour change.

Where we are going

Ageing is universal, disease is not, and it can be considered as a system failure. The human system is built with many fall-back systems. With ageing there is inherent system deterioration but a system of checks and balances can keep this system running optimally, without loss of function. The system needs to be considered holistically, an engine is not merely the sum of its parts, but a problem in one area ultimately determines the mileage the engine will achieve as well as the performance. However, how can this be addressed when a systematic diagnostic tool does not exist? Understanding the key relationships between systems and having a concise operationalised definition of system performance, thus of successful ageing, is required to not only determine when interventions are required, but also how and where they should be implemented, as well as a measure of intervention performance.

If the twentieth century was the century of population growth, then the twenty-first century will become the century of ageing.[27] As a species we have made great strides in life expectancy, however with those gains come challenges.[27] Age-related diseases are mounting as a result of

healthy life expectancy lagging behind increased life expectancy. [27] In the past and maybe the present, health has been over simplified into single units of measure. Given this over simplification there is substantial room for the development of personalised health. This is particularly true in an era when we have the ability to measure everything that matters, from activity trackers, sleep, vital signs, blood pressure, heart rate and stress and develop algorithms in combination with genetic and physiologic information for the purposes of making personalised recommendations.[28] We need to consider the synergisms and interactions between different aspects of human life in a broader sense. Interactions between different components and the complex health relationships between health and disease should not be over simplified into single unrepresentative health indicators. Within the context of such reductionist type approaches, we need to be mindful that the development of such scales should be useful regardless of the perspective of the users. This is also a criticism of Rowe and Kahn's two factor model, as it fails to acknowledge that diverse experiences can lead to different interpretations of the meaning of success.[7] The relevance of this can be illustrated with a basic one to ten pain scale, in which the level of pain described by a patient is strongly dependent on previous experience, tolerance and affliction. Ignorance to any of these, although likely unintentional, could lead to various misdiagnoses and inappropriate medication.

The next stages in healthcare development will try to incorporate large amounts of information and may result in so called deep phenotyping. The sequencing of the human genome opened the door to characterising traits of health and disease and linking it to genetic information.[29] However the function of many genes remains unknown and what is known, is limited to a few

cell types, tissues or physiological contexts.[29] Difficulties in information collection, differences in disease manifestations, descriptive phenotypes or disease subclasses requires extensive examination of the discrete components of disease phenotypes, information which is not typically recorded in medical charts and further complicates elucidation.[29] Delving into this information may help to link seemingly unrelated conditions which share common biological pathways and/or disease mechanisms.[29]

Understanding the origins of disease, as well as the common pathways of common co-existing diseases, not only grants us the opportunity to develop targeted treatments, but it also gives an opportunity to develop personalised preventative interventions. Recent developments in scientific literature suggest it may be possible and realistic to slow the ageing process.[30] Delaying ageing could increase life expectancy by 2.2 years most of which could be spent in relatively good health, while saving $7.1 trillion dollars over the next fifty years according to an American simulation study.[30] The same efforts put into heart disease and cancer treatment would result in declining returns as improvements in health and longevity would diminish by 2060 according the same model.[30] Both of which suggest that efforts to maintain health and prevent disease are highly efficient but also that they should be top research priorities.[30] Moreover, if chronic metabolic diseases were dealt with using appropriate dietary strategies, statistical models estimate the death insurance claims would drop by 13 percent, meaning a reduction in premature loss of life from preventable conditions.[31]

In general, successful ageing should be defined based on insights into the underlying contributors to maintained physical and mental health within the context of possible chronic

disease. This definition considers the ability to maintain physical (independent) functioning for as long as possible. Definitions should support the setting of realistic goals, as well as appropriate new goals in order to reflect the dynamic as well as personal nature of ageing.

What can this thesis contribute?

"We can't look at just one or two phenotypes because we don't know the function of most genes nor can we make assumptions about what to look for."[29]

This thesis aims to further the discussion on the contributing aspects to what it means to age successfully. Here we attempt to explain the past, and predict the future of ageing research by quantifying past research, proposing new models as well as stepping into the future by combining genetic and lifestyle factors for chronic disease development. Lastly, we propose an innovative psychological intervention to aid in maintaining successful ageing.

In *Chapter Two* we delve into the recent past and make recommendations for defining and operationalising successful ageing by performing a systematic review of recently published healthy ageing models. In *Chapter Three* we test a holistic model of successful ageing, the healthy ageing phenotype, using confirmatory factor analysis in order to determine if health should be quantified into traditionally separate '*health branches'* or domains, i.e. physiological wellbeing, social wellbeing, psychological wellbeing, physical capability and cognitive function. In *Chapter four* we attempt to operationalise successful ageing ourselves using a data driven model. We use exploratory factor analysis to determine if it is not only possible to make logical groupings of health parameters (domains) but also if these can be operationalised into scores

which are predictive of meaningful health outcomes including walking speed, self-rated health, emotional vitality and dependency. In *Chapter Five* we perform a systematic review and meta-meta-analysis to explore relevant contributors to COPD as an accelerated ageing syndrome. Here we examined the influence of single nucleotide polymorphisms and (environmental) exposures which have been identified in meta-analyses as being relevant with respect to the risk of developing COPD. In *Chapter Six* we explore the biological context of these genetic alterations in order to understand signals within the data noise. Specifically, we employ network and variant effect predictor analysis in order to determine potentially where, how and if specific gene variations influence COPD susceptibility and/or pathology. In *Chapter Seven* we present the study design of an ongoing randomized clinical trial to investigate the efficacy of working memory training in COPD patients on cognitive performance including self-control that may aid patients in complying to healthy diet and exercise regimes. Finally, in chapter eight we summarise our findings as well as discuss future and ethical implications of advancements in personalised health. We discuss how caution should be taken with the availability of data for prediction models as well as models themselves; without pre-emptive precautionary measures there is significant risk of data misuse.

References

1. Organization WH. World report on ageing and health: World Health Organization 2015.
2. Michel JP, Dreux C, Vacheron A. Healthy ageing: Evidence that improvement is possible at every age. *European Geriatric Medicine* 2016;7(4):298-305. doi: https://doi.org/10.1016/j.eurger.2016.04.014
3. Kusumastuti S, Derks MGM, Tellier S, et al. Successful ageing: A study of the literature using citation network analysis. *Maturitas* 2016;93:4-12. doi: https://doi.org/10.1016/j.maturitas.2016.04.010
4. Jin K. new perspectives on healthy aging: Elsevier, 2017.
5. Young Y, Frick KD, Phelan EA. Can Successful Aging and Chronic Illness Coexist in the Same Individual? A Multidimensional Concept of Successful Aging. *Journal of the American Medical Directors Association* 2009;10(2):87-92. doi: https://doi.org/10.1016/j.jamda.2008.11.003
6. Rowe JW, Kahn RL. Human aging: usual and successful. *Science* 1987;237(4811):143-49.
7. Rowe JW, Kahn RL. Successful aging. *The gerontologist* 1997;37(4):433-40.
8. Stowe JD, Cooney TM. Examining Rowe and Kahn's concept of successful aging: Importance of taking a life course perspective. *The Gerontologist* 2014;55(1):43-50.
9. Brown LJ, Bond MJ. Comparisons of the utility of researcher-defined and participant-defined successful ageing. *Australasian journal on ageing* 2016;35(1):E7-E12.
10. Masoro EJ. "Successful Aging"–Useful or Misleading Concept?: Oxford University Press, 2001.
11. GOLD. Global Strategy for the Diagnosis MaPoC, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2018. (https://goldcoped.org/. Accessed July 17, 2018); 2018.
12. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet* 2012;380(9859):2095-128.
13. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine* 2006;3(11):e442.
14. MacNee W. Is chronic obstructive pulmonary disease an accelerated aging disease? *Annals of the American Thoracic Society* 2016;13(Supplement 5):S429-S37.
15. Barnes PJ. Senescence in COPD and its comorbidities. *Annual review of physiology* 2017;79:517-39.
16. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. *American journal of respiratory and critical care medicine* 2013;188(8):e13-e64.
17. Petty TL. The history of COPD. *International journal of chronic obstructive pulmonary disease* 2006;1(1):3.
18. Organization WH. Noncommunicable diseases country profiles 2018. Geneva, 2018.
19. Health UDo, Services H. National Institute on Aging and World Health Organization (2011). *Healthy workplaces: A WHO global model for action*
20. Kochanek KD, Murphy SL, Xu J, et al. NCHS Data Brief Mortality in the United States. *US Department of Health & Human Services* 2017;293
21. Wouters EF, Wouters BB, Augustin IM, et al. Personalised pulmonary rehabilitation in COPD. *European Respiratory Review* 2018;27(147):170125.
22. Troosters T, Blondeel A, Janssens W, et al. The past, present and future of pulmonary rehabilitation. *Respirology* 2019
23. Tønnesen P. Smoking cessation and COPD: Eur Respiratory Soc, 2013.

24. Spruit MA, Pitta F, McAuley E, et al. Pulmonary rehabilitation and physical activity in patients with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2015;192(8):924-33.
25. Egan C, Deering BM, Blake C, et al. Short term and long term effects of pulmonary rehabilitation on physical activity in COPD. *Respiratory medicine* 2012;106(12):1671-79.
26. Richardson J, Loyola-Sanchez A, Sinclair S, et al. Self-management interventions for chronic disease: a systematic scoping review. *Clinical rehabilitation* 2014;28(11):1067-77.
27. Greaves CJ, Sheppard KE, Abraham C, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. *BMC public health* 2011;11(1):119.
28. Ordovas JM, Ferguson LR, Tai ES, et al. Personalised nutrition and health. *BMJ* 2018;361:bmj. k2173.
29. Delude CM. The details of disease. *Nature* 2015;527(7576):S14.
30. Goldman DP, Cutler D, Rowe JW, et al. Substantial health and economic returns from delayed aging may warrant a new focus for medical research. *Health affairs* 2013;32(10):1698-705.
31. Schoonbee J, Elefteriadis E. Why would a leading global reinsurer be interested in nutrition? *BMJ* 2018;361:k2435.

Chapter 2

Towards a multidimensional healthy ageing phenotype

Authors:

Sarah Mount[a] , Jose Lara[b,c], Annemie M.W.J. Schols[a] , and John C. Mathers[b]

[a] Department of Respiratory Medicine, Maastricht University Medical Centre, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht, The Netherlands,

[b] Human Nutrition Research Centre, Institute of Cellular Medicine and Institute for Ageing, Newcastle University

[c] Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, UK Correspondence to Professor John C. Mathers, Human Nutrition Research Centre, Institute of Cellular Medicine and Institute for Ageing, Newcastle University, Biomedical Research Building, Campus for Ageing and Vitality, Newcastle upon Tyne NE4 5PL, UK. Tel: +44 0 1912081133; e-mail: john.mathers@ncl.ac.uk

Abstract

There is great interest in developing tools to measure healthy ageing and to identify early stages of health impairment which may guide the implementation of interventions to prevent or delay the development of disease, disability and mortality. Here we review the most recent developments directed to operationalise, and test, definitions of healthy ageing.

Recent findings suggest that there is a lack of consensus about how to define healthy ageing and, unsurprisingly, diversity in the instruments for its measurement. However, progress is being made in describing and in devising tools to capture the healthy ageing phenotype. Attempts to measure healthy ageing have relied primarily on cross-sectional data collected in older people. More recent studies have assessed the healthy ageing phenotype using markers of multiple functional domains and have used longitudinal data to model the dynamics and trajectories of healthy ageing.

Given the complexity of the ageing process, no single measure is able to predict the ageing trajectory. Current attempts to operationalise the healthy ageing phenotype have relied on markers and data from earlier cohort studies and are limited by the tools used to collect data in those studies. Such data are often unsuitable to detect early subtle declines in function and/ or are inappropriate for use in younger old adult populations. Future studies employing more objective and novel markers of healthy ageing are likely to offer opportunities to define and operationalise the healthy ageing phenotype.

Introduction

What is ageing? Ageing has been described as the biological changes that occur in an individual that are associated with a gradual decline in function. For most, ageing is experienced as a slow deterioration resulting eventually in frailty, disability, disease and death. However, evidence from model systems and from human studies shows that the ageing process is malleable, the ageing trajectory can be slowed and the link with disease weakened. Given favourable circumstances, individuals can maintain good physical and cognitive function. In part those circumstances are genetic but socioeconomic factors and lifestyle are major determinants. Recent data from the USA illustrate the strong association between higher income (as a surrogate for more favourable circumstances) and longevity.[1] This study also showed that whilst those with higher income gained about 0.2 years of life extra per year over the period 2000 – 2014, the poorest had no improvements. In addition, Chetty and colleagues noted that income was associated strongly with higher physical activity and with lower likelihood for smoking and obesity – lifestyle-related factors which are established modulators of ageing and risk of age-related disease.[1]

Reaching consensus on how healthy ageing should be defined has proven to be a difficult task. This lack of consensus is a significant impediment not only for research but also for national surveillance programmes, public health interventions and for commercial developments. This difficulty is due partly to the different perspectives, expectations and aims of researchers from different disciplines. While from a scientific perspective the preservation of health and function is a priority, evidence indicates that other more subjective aspects of wellbeing are also important to the individual. Interviews with older people have indicated that wellbeing is produced by having the "capability" to mobilise resources to achieve contextually appropriate

goals and to respond effectively to changing personal circumstances.[2] Lastly, there is heterogeneity in the terminology used to denote the concept of healthy ageing in the literature. For the purpose of the present review we will adopt the term healthy ageing.

The importance of defining and measuring healthy ageing is underlined by the fact that the global population is ageing; By 2080, 29% of the European population will be aged > 65 years.[3] This is due to the combination of reduced birth rate and increased life expectancy. To reduce the risk of economic instabilities and to prevent social collapse as a result of too few people in economically productive work, several countries are raising the retirement age to maintain the workforce and to reduce the pension burden associated with greater longevity. In addition, because ageing is the major driver of most common complex diseases, the chronic disease burden is increasing.[4] These pressures emphasise the importance of finding ways to enable people to age better and to maintain good function and high levels of wellbeing. To facilitate the development of targeted interventions which may be more resource efficient, it would be helpful to distinguish those who are ageing well from those who are ageing less well. In other words, we need to be able to identify a healthy ageing phenotype.[5] The benefits of maintaining good health into old age include saving health care costs, improving quality of life, and enabling older people to continue to participate productively in society and to offer their stability, heightened capacity for synthetic problem solving, increased ability to manage conflicts, and ability to consider perspectives from other age groups.[6]

Models of healthy ageing

Cellular dysfunction is the biological basis for the age-related decline in function and for the increasing risk of frailty, disability and disease, the cardinal features of ageing.[7] The accumulated macromolecular damage in ageing is pervasive affecting virtually every cellular, tissue and whole-body function and is remarkably similar in multiple species. This observation underpinned the recent proposal for nine hallmarks of ageing[8] which help to conceptualise and systematise a highly complex collection of processes. Importantly, the ageing process is plastic and the accumulation of molecular damage and cell dysfunction can be slowed.[9] Models of healthy ageing have been based on Rowe and Kahn's 1987 proposal which differentiated between healthy older individuals and those with disease and/or disability,[10 11] Rowe and Kahn's model adopted a multi-dimensional approach to ageing and proposed that healthy ageing is a combination of a low probability of disease and disease-related disability, high cognitive and physical functional capacity, and active engagement with life.[11] More recently there has been emphasis on models which include both subjective e.g. psychosocial wellbeing as well as objective, i.e. biological, measurements.[12] Psychological based models emphasise "how" healthy ageing occurs whereas the biological models emphasise the "what".[6]

The different conceptual frameworks of healthy ageing models have been reviewed. Depp and Jeste performed a meta-analysis on the definitions of successful ageing as well as the attempts to operationalise them.[13] Martinson and Berridge conducted a systematic review to analyse the range of critiques of successful ageing models and the suggestions for improvement from the social gerontology literature.[14] More recently Anton et al. provided an overview which focussed on physical function, and the role of interventions that may enhance mobility and physical function and so promote independence among older adults.[15]

Healthy ageing phenotypes

Currently most literature focuses on morbidity and mortality as ageing phenotypes. Our group, and others, have emphasised the need to focus on a combination of objective and subjective outcomes including physical capability, cognitive function, physiological and metabolic health, and psychosocial well-being,[12] In addition, the importance of assessing sensorial functions has been identified.[16] Olfactory function may be an indicator of the integrity of the ageing brain in older people,[17] since smell dysfunction is among the earliest "preclinical" signs of neurodegenerative diseases such as Alzheimer disease and sporadic Parkinson's disease.[18]

Frailty vs. Healthy ageing

It could be argued that the development of frailty is an example of failure to age healthily. Frailty indicators have been extensively validated in multiple populations for their ability to predict age-related adverse health outcomes ranging from falls to mortality in elderly populations.[19] Measures of frailty have been reviewed recently by Roppolo et al. (2015) who showed that different instruments captured different characteristics of frailty and that whether an individual is classified as frail or not depends on the index of frailty used in the assessment.[20] Whilst frailty may be evidence of a failure to age healthily, it is unlikely that assessments which are used to assess frailty would be sufficient to assess healthy ageing. This is because the instruments used in such assessments are designed for individuals experiencing a substantial degree of disability or illness and are usually applicable only to the oldest segments of the population. Because of floor and ceiling problems, such tools would be insensitive when applied to younger individuals. However, recently, Romero-Oturo (2013) proposed that a simple 5-item index called the Frailty Instrument (FIt) for primary care which

is based on the *Survey of Health, Ageing and Retirement in Europe* (SHARE) may be useful for assessing and monitoring frailty in community dwelling people over the age of 50.

Current evidence on operationalisation of definitions of healthy ageing

To identify recent studies operationalising healthy ageing, a systematic search was performed using PubMed from 01-01-2013 until 01-05-2016. Publications were included if the authors intended to measure the healthy ageing phenotype or employed a multidimensional approach to measure healthy ageing. Using this strategy, twelve studies were found (Table 1). All tools which are included in this brief review were multidimensional.

Although evidence is still scarce, we identified several studies that focused on operationalising healthy ageing. Among the papers identified, the very recent paper by Tampubolon[21] is novel in using the concept of the healthy ageing phenotype[12] and in assessing the trajectories of some of the biomarkers proposed in that model among participants of the English Longitudinal Study of Ageing (ELSA). ELSA is a prospective, nationally-representative sample of people aged ≥ 50 years. The biomarkers included measures of cardiovascular function, glucose homeostasis, lung function, adiposity, lipid metabolism, and inflammation. This study showed a secular decline in healthy ageing from middle-age, which was sharper among women and differed by socioeconomic position. Those with greater material advantage and higher educational attainment had smaller declines in biomarkers of healthy ageing.[21] In addition, the findings were in line with previous studies (e.g. Pruchno et.al.) suggesting that maintenance of healthy ageing at follow up was more common among men, those who were working, the married, and those having better social relationships at baseline.[22]

Two other studies proposed different tools to assess healthy ageing.[23 24] Tyrovolas et al., used a composite index of healthy ageing which included 10 elements. education, financial status, social activities (subdivided into friends, family and excursions per year), CVD risk, BMI, depression, physical activity and Mediterranean diet. Using data on older adults from Mediterranean countries, these authors reported that a 1-point increase in their 10-point index was associated with one less annual visit to health centres. Using data from the Medical Research Council's Cognitive Function and Aging Study (CFAS), Cosco et al. tested an ageing index based on activities on daily living, cognitive function, and subjective aspects such as personal resources and engagement.[24] These authors reported that their index was associated with use of health services, informal care and other services such meals on wheels.

The Whitehall study, a longitudinal cohort of civil servants in the UK with a long follow-up period (median 16 years) and a large sample size, has been recently used to classify individuals as successfully ageing or not. In this study, healthy ageing has been defined as survival in the follow up period (mean age 60), with no diagnosis of chronic disease or abnormal oral glucose tolerance test, no mental health problems and normal cognitive (Alice Heim 4-I, short-term verbal memory test, 2 tests of verbal fluency, Mill Hill Vocabulary test), cardio metabolic (SBP), respiratory (FEV1/height^2 in L/m^2) and musculoskeletal function (walking speed over a clearly marked 8-foot walking course).[25]

Discussion

To date, few studies have attempted to operationalise healthy ageing using comprehensive and multi-dimensional approaches (such as that proposed by Lara et al. 2013[12]) and to apply

the resulting tools to data from current longitudinal cohorts. Current studies have relied on data with only partial or proxy measures for each of the different domains associated with the healthy ageing phenotype. The available indices focus on different combinations of domains of healthy ageing but commonly focus on cognitive function and measures of physical capability. Decline in cognitive function is a hallmark of ageing[26] and recent data from the Sydney Memory and Ageing Study (an observational population-based cohort study) showed that both baseline cognitive ability and decline in cognitive ability over two years predicted mortality, even in the absence of dementia.[27] Cognitive decline also has important financial, personal and societal consequences, and is the cause of 40% of admissions to institutionalised care in the UK [26]. Most of the studies reviewed included at least one measure of global cognitive functioning, commonly the mini mental state examination (MMSE). This tool has been used widely in, and accepted as appropriate for, elderly populations. However, its ability to detect subtle deficiencies, namely mild cognitive impairment,{Delude, 2015 #49} or cognitive changes at high levels of cognition such as among highly educated people[29 30] has been questioned. A recent Cochrane review concluded that the MMSE was insufficient as a stand-alone single-administration test in the identification of MCI patients who could develop dementia.[31] Subtle changes in cognitive function can be meaningful and the majority of tools ignore this by using tools designed to identify more evident declines in cognitive function. Cullen and colleagues identified 39 screening tools, found that most large-scale community screening programmes used informant rated scales which could be carried out by telephone or post and noted that it is likely that "one size does not fit all" in cognitive screening.{Delude, 2015 #49}

The reviewed studies have mostly used activities of daily living, with few using more objective measures of gait or walking tests. All but one study focused on older populations, those aged 70+ and often in 85 to 90+ individuals. Puchno et al. found that midlife predictors of healthy ageing differed from those which apply later in life.[22] In addition, it appears that some biomarkers of ageing which appear robust in younger old individuals may not be valid in very old people. Indeed, in some cases the reverse may apply e.g. higher BP is a risk factor in younger people but may be a protective factor in very old people.[32] Overall, assessment of the utility of markers of healthy ageing is limited by uncertainties and lack of consensus about the appropriate outcome measures to be used in such assessments. Whilst earlier studies have focussed on hard end points such as death or diagnosis of major age-related disease, some recent studies have considered other outcomes such as use of health services (e.g. hospitalisations).

None of the studies identified in the current search have used markers of sensorial functions. Sensory functions are critical for normal function, independence and social interaction and most decline with age.[16] Smell dysfunction is one of the earliest indications of preclinical neurodegenerative diseases[18] but the predictive value of sensory function for age-related health outcomes has yet to be validated.[16] Lastly, few tools consider social function in the assessment of the healthy ageing trajectory despite the fact that social interactions and personal support networks are strongly associated with both morbidity and mortality.[33-35]

Conclusion

Healthy ageing is the maintenance of a range of functions including physical capability, cognition, physiology, social, sensory and psychological wellbeing. A suitable approach to measure healthy ageing should include both subjective and objective assessments of as many as possible of these factors since the complex interplay of these factors determines health and wellbeing outcomes.[15] However, this must be balanced against the resource requirements for more comprehensive assessments and future research should aim to identify the minimum set of measures which provides reliable prediction of the ageing trajectory and which could be used as outcome measures for interventions.[12] However, current models of healthy ageing remain incomplete; the operationalisation of the healthy ageing phenotype is a work in progress.

Key Points

- The lack of an agreed definition of healthy ageing limits the development of tools for its measurements and, therefore, the ability to predict the ageing trajectory.
- Current models focus on absence of frailty and on functional status including the ability to carry out the activities of daily living.
- Conceptualisation and operationalization of the healthy ageing phenotype offers a potential route to the development of tools for assessing ageing which are age-, sex- and culturally-appropriate.
- Tools for healthy ageing will need to be validated in younger and older populations in a range of setting to determine their generalizability.
- Testing of tools for healthy ageing is limited by uncertainties about the most appropriate outcomes measures to use at different stages in the life-course. Outcome

measures such as hospitalisation, institutionalisation and disability which are sensitive in younger populations need to be developed and validated.

Table 1

Name	Format	Format of study	Characteristics of tool	Outcome parameters	Type of analysis	Domains included	Variables	Age	Reference
Healthy Ageing Index	Index	Cohort study; median follow-up 12.8 years	Items scored in tertiles, 0 being the healthiest and 2 being the unhealthiest; glucose treated as a continuous variable (decimal cut offs applied); Where gender differences exist different cut offs for men and women were applied;	Index score from 0 to 10; predicts mortality and morbidity	Cox proportional hazards model; Mortality - optimised HAI (applied regression coefficients from survival models such that the index ranges from 0 to 10; strongest associated components given stronger weights	Physiological, psychological	sbp, fvc, mmse, sCrt, Glc; Covariates: gender, age, smoking history, BMI, education, CHD, cerebrovascular disease, diabetes, pulmonary disease, kidney disease, osteoarthritis, and depression used as covariates, race, PA, depression (CES-D),	> 65	[36];
Genetics of Healthy Ageing	Survival Prediction Model	Cohort study; median follow-up 6 years	Hazard rations have been calculated for individual variables	Mortality hazard ratios	Multi-variate cox regression model for survival predictors	Physiological, Physical capability, Psychological, Behavioural, Cognition	SES (marital status education, occupation), ADL, sensory functioning, sMMSE, lifestyle (smoking, drinking habits), health and morbidity (present and past), perceived health, medications, haematological parameters: (Crt, GLc, ALT, TChol, HDL, LDL, trigly), hospitalisations, weight loss in past year, psychological wellbeing, height weight, physical functioning (ability to see person at 4 meters without aid, ability to hear without aid, ability to go outside, up and down stairs, ability to exercise unassisted; handgrip strength) DNA,	90+	[37]

Successful Ageing Index	Index	Cohort study, 2 year follow-up	Constructed model from lay perspectives of healthy ageing (personal resources; optimism, personal engagement (interest in life), loneliness and self-awareness), modified Katz model of activities of daily living, and cognitive functioning (MMSE). Ordinal items were assigned values based on the scores of each of the times used to create a continuous value. Depending on the number of possible answers, fraction values were assigned accordingly	Use of health care services	Logistic regression model	Physiological, Psychological, Social	ADL (dressing, going upstairs, continence, cutting toenails), IADL (doing heavy housework, shopping, preparing meals, reaching overhead shelf, tying knot, housekeeping, getting a bus, managing finances), MMSE, SES (education, age), gender,	>65	[24]
Successful ageing index	Score	Cohort study; median follow-up 6 years	Cumulative Score 0-10; Each aspect of the model given scores from 0 to 10, positive or negative according to their influence on health; model adjusted for confounders; 1/10 unit increase in index associated with a 0.8 less annual visits to healthcare centres; stratified analysis for gender, revealed heterogeneity in the predictors for successful ageing between men and women	Use of health care services	Principle component analysis used to verify multi/ dimensionality of data. Nested multiple linear regression models to evaluate the association between number of annual visits to health care centres (dependent outcome) and participants' characteristics, and their level of successful ageing (independent variables)	Behaviour, Physiological, Psychological, Social	Model: SES (education, financial status), depression, participation in social activities (social with friends, social with family, excursion participation), CVD risk (Glc, SBP, HDL, LDL, Trigly, WC,), BMI, Lifestyle (PA), MEdDietScore, depression; Covariates: smoking status, living alone, gender, age	65 - 100	[23]

Successful Ageing Index	Index	cross-sectional	4 dimensional model: physical and functional health, psychological welling and cognition , social engagement and family support, economic resources and financial security ;(each dimension had 2 indicators with equal weighting); for each indicator, if the participant demonstrated wellbeing in that domain they received one point, if they did not they were given no points; score ranged from 0-8 (8 = all rounded SA)	interviewer-rated health	Construct validity of SAI was examined for its association with interviewer rated health in a multiple binary logistic regression model; correlates of SAI which were investigated using a multi-regression model: biomarkers of physical health: handgrip strength,	Physiological, Psychological, Social, Cognition	Physical health: self-rated heath, ADL, (good health = good self-rated health+ independence in activities of daily living); Psychological wellbeing: PANAS, MMSE; Social engagement and support; economic resources; correlates: personality (optimism), socio-environmental (barriers to social activities), demographic (living arrangement, education attainment), full blood cell count, kidney and liver function, c-reactive protein, HBA1c, HDL, LDL TChol, trigly, and albumin, gender	>85	[38]
Cognitive abilities screening instrument	Multi-index model	Cohort study; 3 and 6 year follow up	multi-index model; frailty index, social vulnerability index, protective factors index combined to predict changes in cognitive function	Cognitive function (worsening or improvement); mortality	Poisson and logistic regression models to predict cognitive function; each index modelled separately to analyse the effects on the cognitive function, corrected for age, gender, education frailty, social vulnerability, protective factors. Produced a multi-index model, examined models for both 3 and 6 year follow up; all variables multiplied by 10 such that the results in the betas represent 10% increases in the beta	Cognitive function, physical capability, Social, behaviour	Cognition: attention, concentration, orientation, short term memory, long term memory, language abilities, visual construction, list generating, fluency, abstraction, and judgement. Each of the following are dichotomised (absent, present; given a 0.5 score per deficit) Frailty index : accumulation of deficits; Social vulnerability index: (social network, marital status, living situation); Protective factors index PA, use of antihypertensive medication, no smoking,	71-93	[39]

							moderate alcohol consumption, good self-rated health, healthy weight		
Successful ageing	Successful ager classification	Cohort study;16-year follow-up	Successful ageing categorisation: being free of major chronic diseases and having good physical, cognitive, respiratory, cardiovascular, and mental health	Mortality and morbidity predictions associated with (un)successful ageing classifications	Logistic regression to determine associations of baseline BMI and WC categories with survival and successful ageing	Physiological, cognition, psychological, Physical capability	Good functional status was defined as not being in the worst quintile of any of the domains assessed; Mental health score (SF-36), disease status, global cognitive score, walking speed, FEV1, WC; Covariates: age, sex, ethnicity (white, south Asian, black, other), education, smoking status, obesity. The Framingham general cardiovascular disease (CVD) risk score (age, HDL, LDL, TotChol, SBP, smoking, and diabetes), PA	35-55	[40]; Used in many publications
Successful Ageing profile	Profile	cross-sectional	Classified into successful/unsuccessful profiles	Classification verification	Cluster analysis to identify distinct patterns of successful ageing in centenarians; groups individuals into homogenous subsets. Logistic regression used to verify predictors of successful ageing considering the cluster structures. Exploratory analysis used to examine model fit	Cognition, psychological, physiological, social	Shortened MMSE, Depression (GDP), social engagement frequency, religion, health status, IADL, morbidity (hypertension, diabetes etc.), ADL, satisfaction with life scale; Covariates considered: gender, marital status, education, living arrangements, psychological resources (self-efficacy, purpose, hope), futurity persistence), and social and economic resources (number of living children, satisfaction with social support, income per month, and	<100	[41]

							income adequacy for expenses)		
Successful ager	Successful ager	cohort study; 4-year follow-up	Successful ager classification based on a combination of objective and subjective markers	Change in classification	Latent profile analysis used to identify successful agers from unsuccessful agers based on the subjective and objective criteria. Multinomial logistic regression used to determine the effect of influential variables on successful ageing status	Physical capability, Physiological	Objective measures: diagnosis of arthritis, hypertension, CVD, cancer, diabetes, osteoporosis, stroke, and lung conditions, functional abilities (walking for a quarter of a mile, walking up 10 steps, standing for 2 hr, stooping), pain assessment; Subjective measures: ratings of ageing success and quality of life; Early life covariates: Age, race, marital status, incarceration, number of children, gender, education; Midlife covariates: marital status, employment, volunteer, smoking status, alcohol consumption, BMI, social support, religiosity, exercise	50-74	[22]

Mortality risk	hazards ratio	Cohort study: 4,9-year follow-up	Mortality risk calculation based on objective assessments of physical capability and cognitive function	5-year mortality risk predictor for females	Data compared across three mobility phenotypes and across three cognition phenotypes using chi-square tests (categorical variables) and analysis of variance (continuous variables). Kaplan–Meier estimates of survival functions used to depict survival by mobility phenotype stratified by cognition phenotype. Associations of combined mobility–cognition phenotypes (nine combinations), mobility phenotypes, and cognition phenotypes with risk of mortality were analysed using Cox proportional hazards models.	Physical capability, cognitive function,	Short Physical Performance Battery; Trails B, 3MS, a 100-point extended version MMSE, CVLT, Digit Span; category, verbal fluency tests. Covariates: BMI, ethnicity, educational self-reported health, hospitalisation in the past year, smoking status, depression, exercise and ability to perform basic activities of daily living, diagnosis of myocardial infarction, stroke, congestive heart failure, hip fracture, diabetes, arthritis, Parkinsonism, COPD, and cancer excluding non-melanoma skin cancer. A comorbidity score was calculated as the sum of these comorbid conditions (range 0–9).	65+	[42]
Health related quality of life measure	Equation	16-week intervention study	Health relate quality of life measurement tool used as a proxy to model declines in health	Health related quality of life	HRQOL domains (physical, mental, and social) quantified by logistic growth equation HRQOL quantitative model consists of one equation for each health domain (average level between the self-report health status and experienced-health components). Equations specifies how domains change as a function of itself and the other	Physical capability, psychological, Social well being	The Short Form 36 Health Survey, the Lubben Social Network Scale-6, Friendship Quality Scale, employment status, education, MMSE, sex, age, marital status	65-90	[43]

					components in the system.				
Healthy Ageing Phenotype	Score	Longitudinal Study (9 years)	Score was calculated using tertiles or clinical cut-offs (coding 0, 1 and 2 accordingly) of each of the eight biomarkers, then added the codes to give a score which ranges from 0 to 16.	Score 0-16	Linear mixed model with random intercepts was used to estimate trajectories of change	Physiological, Others included as covariates behavioural, psychological	SBP, Glc, HbA1C, FEV-1, WC, HDL, Trigly, CRP; Covariates: gender, age, social class (, wealth, education), marital status, comorbidities (CVD, COPD; diabetes; stroke; arthritis; osteoporosis, cancer, depression), Behavioural (smoking), drinking (and physical activity	≥50	[21]

Abbreviations

SBP systolic blood pressure;

FVC forced vital capacity;

FEV-1 forced expiratory volume in one second;

MMSE- Mini-mental state examination;

sCRT serum creatinine, Glc glucose;

ADL Activities of daily living;

TChol total cholesterol;

HDL High density lipoprotein;

ALT Alanine aminotransferase;

CRT creatinine;

Trigly Triglycerides;

IADL instrumental activities of daily living;

sMMSE standardised MMSE;

PA Physical activity;

WC waist circumference;

MedDietScore Mediterranean diet;

PANAS Positive and Negative affect schedule;

HbA1c Glycated haemoglobin;

FEV1 Forced expiratory volume in one second/height2 in L/m2;

CBVD cerebrovascular disease;

HGS handgrip strength;

CVD cardiovascular disease,

COPD chronic obstructive pulmonary disease.

Declarations

Acknowledgements

None.

Financial Support

This work was supported by the Maastricht University Interfaculty Programme 'Eatwell'.

Conflict of Interest

None.

References

1. Chetty R, Stepner M, Abraham S, et al. The association between income and life expectancy in the United States, 2001-2014. JAMA 2016;315(16):1750-66.
2. Heaven B, O'Brien N, Evans EH, et al. Mobilizing resources for well-being: implications for developing interventions in the retirement transition. The Gerontologist 2015:gnu159.
3. EuroStat. EuroStat: Statistics Explained Luxembourg: European Commision; 2015 [updated 7 December 2015. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing#Past_and_future_population_ageing_trends_in_the_EU accessed December 15 2015.
4. Murray CJL, Barber RM, Foreman KJ, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. The Lancet 2015;386(10009):2145-91.
5. Franco OH, Karnik K, Osborne G, et al. Changing course in ageing research: the healthy ageing phenotype. Maturitas 2009;63(1):13-19.
6. Rowe JW, Kahn RL. Successful Aging 2.0: Conceptual Expansions for the 21st Century. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 2015;70(4):593-96. doi: 10.1093/geronb/gbv025
7. Kirkwood TBL. A systematic look at an old problem. Nature 2008;451(7179):644-47.
8. López-Otín C, Blasco MA, Partridge L, et al. The hallmarks of aging. Cell 2013;153(6):1194-217.
9. Newman AB, Glynn NW, Taylor CA, et al. Health and function of participants in the Long Life Family Study: a comparison with other cohorts. Aging (Albany NY) 2011;3(1):63-76.
10. Rowe JW, Kahn RL. Human aging: usual and successful. Science 1987;237(4811):143-49.
11. Rowe JW, Kahn RL. Successful Aging. The Gerontologist 1997;37(4):433-40. doi: 10.1093/geront/37.4.433
12. Lara J, Godfrey A, Evans E, et al. Towards measurement of the Healthy Ageing Phenotype in lifestyle-based intervention studies. Maturitas 2013;76(2):189-99. doi: http://dx.doi.org/10.1016/j.maturitas.2013.07.007
13. Depp. Definitions and Predictors of Successful Aging: A Comprehensive Review of Larger Quantitative Studies. American Journal of Geriatric Psychiatry 2006;14(1):6-20.
14. Martinson M, Berridge C. Successful Aging and Its Discontents: A Systematic Review of the Social Gerontology Literature. The Gerontologist 2014 doi: 10.1093/geront/gnu037
15. Anton SD, Woods AJ, Ashizawa T, et al. Successful aging: Advancing the science of physical independence in older adults. Ageing Research Reviews 2015;24, Part B:304-27. doi: http://dx.doi.org/10.1016/j.arr.2015.09.005
16. Lara J, Cooper R, Nissan J, et al. A proposed panel of biomarkers of healthy ageing. BMC Medicine 2015;13(1):222.
17. Dulay MF, Murphy C. Olfactory acuity and cognitive function converge in older adulthood: Support for the common cause hypothesis. Psychology and aging 2002;17(3):392.
18. Doty RL. Olfactory dysfunction in Parkinson disease. Nature Reviews Neurology 2012;8(6):329-39.
19. O'Caoimh R, Cornally N, Weathers E, et al. Risk prediction in the community: A systematic review of case-finding instruments that predict adverse healthcare outcomes in community-dwelling older adults. Maturitas 2015;82(1):3-21. doi: http://dx.doi.org/10.1016/j.maturitas.2015.03.009

20. Roppolo M, Mulasso A, Gobbens RJ, et al. A comparison between uni- and multidimensional frailty measures: prevalence, functional status, and relationships with disability. Clinical Interventions in Aging 2015;10:1669-78. doi: 10.2147/CIA.S92328
21. Tampubolon G. Trajectories of the healthy ageing phenotype among middle-aged and older Britons, 2004–2013. Maturitas 2016;88:9-15.
22. Pruchno RA, Wilson-Genderson M. A Longitudinal Examination of the Effects of Early Influences and Midlife Characteristics on Successful Aging. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 2015;70(6):850-59. doi: 10.1093/geronb/gbu046
23. Tyrovolas S, Haro JM, Mariolis A, et al. Successful aging, dietary habits and health status of elderly individuals: A k-dimensional approach within the multi-national MEDIS study. Experimental Gerontology 2014;60(0):57-63. doi: http://dx.doi.org/10.1016/j.exger.2014.09.010
24. Cosco TD, Stephan BCM, Brayne C. Validation of an a priori, index model of successful aging in a population-based cohort study: the successful aging index. International Psychogeriatrics 2015;27(12):1971-77. doi: doi:10.1017/S1041610215000708
25. Sabia S, Singh-Manoux A, Hagger-Johnson G, et al. Influence of individual and combined healthy behaviours on successful aging. Canadian Medical Association Journal 2012;184(18):1985-92. doi: 10.1503/cmaj.121080
26. Deary IJ, Corley J, Gow AJ, et al. Age-associated cognitive decline. British Medical Bulletin 2009;92(1):135-52. doi: 10.1093/bmb/ldp033
27. Connors MH, Sachdev PS, Kochan NA, et al. Cognition and mortality in older people: the Sydney Memory and Ageing Study. Age and Ageing 2015;44(6):1049-54. doi: 10.1093/ageing/afv139
28. Arevalo-Rodriguez I, Smailagic N, Roqué i Figuls M, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). The Cochrane Library 2015
29. Proust-Lima C, Amieva H, Dartigues J-F, et al. Sensitivity of Four Psychometric Tests to Measure Cognitive Changes in Brain Aging-Population–based Studies. American Journal of Epidemiology 2007;165(3):344-50. doi: 10.1093/aje/kwk017
30. Lonie JA, Tierney KM, Ebmeier KP. Screening for mild cognitive impairment: a systematic review. International Journal of Geriatric Psychiatry 2009;24(9):902-15. doi: 10.1002/gps.2208
31. Cullen B, O'Neill B, Evans JJ, et al. A review of screening tests for cognitive impairment. Journal of Neurology, Neurosurgery & Psychiatry 2007;78(8):790-99.
32. Martin-Ruiz C, Jagger C, Kingston A, et al. Assessment of a large panel of candidate biomarkers of ageing in the Newcastle 85+ study. Mechanisms of ageing and development 2011;132(10):496-502.
33. Reblin M, Uchino BN. Social and Emotional Support and its Implication for Health. Current opinion in psychiatry 2008;21(2):201-05. doi: 10.1097/YCO.0b013e3282f3ad89
34. Lyyra T-M, Heikkinen R-L. Perceived Social Support and Mortality in Older People. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 2006;61(3):S147-S52.

Chapter 3

Validation of the healthy ageing phenotype

Authors:

S. Mount [a], A Wesselius[B], C. van der Grinten[c], H. Savelberg[D], T. Van Geel[E], N. Schaper[F], P. Reijven,[G] P. Dagniele[H], M. van Boxtel[I], F. Verhey[J], M. Scham[K], H. Bosma[L], N. Dukers- Muijrers [M], C. Stehouwer [H], A. Schols [a], M. Zeegers[B]

[a] Department of Respiratory Medicine, Maastricht University Medical Centre+, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht, The Netherlands,

[B] Department of Complex Genetics, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+ (MUMC+), Maastricht, the Netherlands.
[C] Department of Respiratory Medicine, Maastricht University Medical Centre, Maastricht, The Netherlands.
[D] Department Human Biology and Human Movement Sciences, NUTRIM School for Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands.
[E] Department of Family Medicine, Maastricht University, Research School CAPHRI, Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands.
[F] Department Internal Medicine, Division Endocrinology, Maastricht University Medical Centre+, Maastricht, The Netherlands.
G Department Clinical Dietetics, University Hospital Maastricht, Maastricht, The Netherlands
[H] Cardiovascular Research Institute Maastricht, Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands.
[I] Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre+, Maastricht, The Netherlands.
[J] Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre+, School for Mental Health and Neuroscience, Alzheimer Centrum Limburg, Maastricht, the Netherlands.
[K] Department of Medicine, Maastricht University Medical Centre+, Maastricht, The Netherlands
[L] School for Public Health and Primary Care-CAPHRI, Department of Social Medicine, Maastricht University Medical Centre+, Maastricht, the Netherlands
[M] Department of Medical Microbiology, Maastricht University Medical Centre+, CAPHRI, Maastricht University, Maastricht, The Netherlands

Abstract

Ageing is a complex process with no single measure able to give an indication of successful ageing. In this study we test whether the *Healthy ageing phenotype* (HAP), based on systematic review of literature and expert opinion, is an appropriate tool for a multidimensional representation of health. In order to determine if the HAP is an appropriate tool, we used cross-sectional data from the Maastricht Study (TMS). Data from 1391 participants, aged between 40 and 75, of which 744 (54%) were female and 312 (23%) were diabetic was available for this purpose. Within this data set, thirty variables were selected based on the HAP's five domain model (i.e. cognitive function, social wellbeing, physical capability, psychological wellbeing, and physiological and metabolic health). Both confirmatory and exploratory analysis were performed on standardised variables. Confirmatory factor analysis indicated poor fit for the proposed five domain model (CFI 0.533 and TLI 0.487). Subsequent exploratory analysis showed a two-domain model, Domain 1 physical, cognitive and metabolic health and domain 2, psychological resilience, including physiological, cognitive, social, and physical capability in the first domain and psychological variables in the second domain. Unfortunately, the theoretical HAP model was not supported in data from TMS. Therefore, new attempts are needed to operationalise successful ageing into a meaningful testable measurement.

Introduction

Global ageing is an emerging health concern.[1] Ageing itself is a complex but universal process which can result in multi-organ functional decline, disability and disease.[2] However, the rate at which this occurs is highly variable and determined by a multitude of factors. Measuring ageing success has proven difficult, and given the complexity of ageing, no single measure is likely to be a good indicator.[2] However, translating research into healthy ageing practice is becoming a ubiquitous challenge due to the forthcoming global burden of aged individuals.[3] Therefore, we need a panel, as well as a way of interpreting measures which capture the key features of ageing success; "*the process of developing and maintaining the functional ability that enables well-being in older age"*.[4] This will allow caregivers and researchers to aid individuals in ageing successfully, as well as to determine the contributors to successful ageing within highly heterogeneous populations. Moreover, it will also aid in designing and evaluating intervention trials effectively and help to ensure the ongoing healthcare and nutrition personalisation trend.

The Healthy Ageing Phenotype (HAP) is such a panel, intended to encapsulate the ability to maintain social relationships, to function independently at both cognitive and physical levels, and to continue to be a productive member of society.[5] Unlike most other tools intended to measure and thus operationalise ageing success, the HAP proposes a panel of biomarkers which comprise a set of functional surrogate endpoints that are influenced by the ageing process, lifestyle interventions and are not limited to specific disease endpoints or mortality.[5] However, surprisingly, no one has tested whether the HAP's compartmentalised operationalisation representation of ageing (five domain structure: social wellbeing, psychological wellbeing, cognitive function, physical capability and physiological and metabolic

health) dynamics is feasible. Therefore, this study aims to determine whether or not the HAP is indeed an appropriate model for representing health in a middle-aged population.

Methods

Participants

Data were obtained from the Maastricht Study (TMS), an observational prospective population-based cohort study.[6] The rationale and methodology have been described elsewhere.[6] In brief, the study focuses on the aetiology, pathophysiology, complications and comorbidities of type 2 diabetes mellitus (T2DM) and is characterised by an extensive phenotyping approach. Eligible for participation were all individuals aged between 40 and 75 years and living in the southern part of the Netherlands. Participants were recruited through mass media campaigns and from the municipal registries and the regional Diabetes Patient Registry via mailings. Recruitment was stratified according to known T2DM status, with an oversampling of individuals with T2DM, for reasons of statistical efficiency.[6]

For the present study, data was available from used 1391 participants, aged between 40 and 75 (mean 59.7, s.d. 8.36), of which 744 (54%) were female and 312 (23%), had a diagnosis of T2DM.

Variable selection

Variables selected for the CFA were similar or comparable to those defined in the key domains of the HAP (table 1), (i.e. (1) physiological and metabolic health, (2) physical capability, (3) cognitive function, (4) psychological wellbeing, and (5) social wellbeing). An explanation of the different variables, including calculations and references can be found in the supplementary materials.

Normality testing

At first, variables were standardised, by division by the range of a variable.[7] Multivariate normality was tested by the use of the Doornik-Hansen test.[8] To accommodate for any multivariate non-normality, the Huber-White scaling correction was applied.[9 10] Missing data was subjected to listwise deletion and were assumed to be missing at random.

Confirmatory Factor analysis

To determine if the hypothesised HAP model exists within the TMS data set, confirmatory factor analysis was conducted, including all selected variables. Model fit was evaluated using comparative fit index (CFI) and the Tucker-Lewis index (TLI), which assumes values between 0 and 1, the latter indicating exact fit. In both the CFI and TLI values above 0.95 are considered indicative of good model fit.[11] Confirmatory factor analysis was performed in R using the Lavaan package.[12 13]

Exploratory factor analysis

Given our findings from the CFA indicated that the hypothesised HAP model does not fit the data from TMS, exploratory analysis was performed on the data set. For this, we first investigated multivariate normality as previously explained (refer to methods). Factorability of the data set was then examined using an anti-image correlation matrix.[14 15] Values of 0.3 or higher were considered appropriate to be included in the EFA. During the EFA it was discovered that the data set was not factorable. As a result, the model was simplified by the removal of offending variables (that is: grip strength, waist circumference, hip circumference, fasting glucose, serum total cholesterol, stroop interference test, practical support, PHQ-9 score).

The remaining variables/dataset, including 1409 observation, was used for the EFA. For this, we used the principle axis factoring method with standardised variables. Variables were standardised by division by the range of a variable.[7] If variables had a loading of (-) 0.3 or higher, on at least one factor they were retained in the model.[16 17] If variables loaded on more than 1 factor, they were assigned to the factor on which they had the highest factor loading.[18] Using the Kaiser criterion (eigen value greater than 1) as well as visual inspection of a scree plot [17], it was determined that the data had a two domain structure. Both promax and varimax rotations were then used in order to determine if the latent factors were correlated.[17] Factors which consisted of at least three variables were considered stable.[17] Finally, to determine the sampling adequacy of the dataset, the Kaiser-Meyer-Olkin (KMO) was applied, as a rule of thumb this value should be greater than 0.6.[19] The resulting two-factor structure is presented in table 2.

Results

The CFA model converged normally after 379 iterations and the robust Comparative Fit Index robust Tucker-Lewis Indexes were 0.533 and 0.487 respectively suggesting that the hypothesised model does not exist in the TMS cohort. The additional exploratory factor analysis, resulted in a two-domain model (table 2). Domain one comprised haemoglobin A1c (HbA1)c, high density lipoprotein (HDL), pulse pressure, body mass index (BMI), Mini mental state examination (MMSE), Groninger Intelligentie Test 2 (GIT), 15-word list delayed recall score (WLTR), processing speed, executive function, emotional support, contact frequency, Chair stand time (TCST) time, gait speed (WT). Domain two comprised generalised anxiety disorder score (GAD), Aggression, personal mastery, General self-efficacy score (GSES) persistence, GSES initiative.

Table 1

Physiological wellbeing	Physical capability	Psychological wellbeing	Social wellbeing	Cognitive function
BMI (kg/m2) (0.11)	Gait speed 6 MWT (m/s) (0.11)	Mean personal mastery score (0.13)	Emotional support score (0.24)	MMSE score (0.04)
Waist circumference (cm) (0.13)	Grip strength (0.04)	General self-efficacy score persistence (0.07)	Practical support score (0.22)	Strooptest interference time (sec) (-0.03)
Hip circumference (cm) (0.09)	Chair stand time (sec) (-0.02)	General self-efficacy score initiative (0.11)	Number of contacts within half a year, divided by 10 (0.07)	Processing speed (0.07)
LDL (mmol/l) (-0.02)		General self-efficacy score effort (0.14)		Executive attention (0.03)
HDL (mmol/l) (-0.05)		Aggression score (-0.07)		WLTR score (0.17)
Serum total cholesterol (mmol/l) (-0.03)		PHQ-9 score (-0.07)		WLT score (0.15)
FEV1 percent predicted (-0.02)		GAD-7 score (-0.09)		GIT score (0.04)
HbA1c (mol) (0.03)				
Fasting plasma glucose (mmol/l) (0.02)				
SBP (0.30)				
DBP (0.04)				

Confirmatory factor analysis variables and hypothesised domains. Here we show the proposed HAP model using variables available with the TMS. Each of the five domains is associated with variables (listed below them), and an associated factor loading. Factor loadings can be interpreted like regression coefficients. The first parameter says that for each unit increase in the latent physiological function (i.e. for each 1SD increase), the model predicts a .11-unit increase in BMI. However, the poor model fit (Comparative Fit Index and the robust Tucker-Lewis Indexes were 0.533 and 0.487 respectively) suggests that this model is invalid for this data set.

Table 2

Variable	Factor1	Factor2	Uniqueness
HbA1c (mol)	-0.4223		0.8003
LDL (mmol/l)			0.9535
HDL (mmol/l)			0.8974
Pulse pressure	-0.3456		0.8334
FEV1 percent predicted			0.9057
BMI (kg/m2)	-0.3710		0.8467
GAD-7 score		-0.5011	0.7224
MMSE score	0.4284		0.8061
GIT score	0.3656		0.8663
WLTR score	0.4150		0.7999
Processing speed	0.6162		0.5959
Executive function	0.5454		0.6931
Emotional support	0.4783		0.7549
Number of contacts within half a year, divided by 10	0.3092		0.8907
Chair stand time (sec)	-0.3802		0.8548
Gait speed 6 MWT (m/s)	0.5740		0.6697
Aggression		-0.3654	0.7781
Mean personal mastery score	0.3035	0.6263	0.5157
General self-efficacy score persistence		0.4524	0.7750
General self-efficacy score initiative	0.3555	0.4545	0.6670

Exploratory factor analysis loadings. Here we show that the variables from the dataset load on to two factors.

Discussion

In this study we aimed to assess and validate the HAP model proposed by Lara et. al. (ref) among individuals from TMS using CFA. Results suggest that the proposed model does not exist, i.e. that the measurements in TMS do not operationalise into the proposed five key

domains of the HAP model, namely social wellbeing, psychological wellbeing, cognitive function, physical capability and physiological and metabolic health. This result was somewhat surprising given the known underlying relationships between the measurements included in these domains, such as those between measures of body composition, BMI, waist, and hip circumference and physiological parameters of fasting blood glucose and blood lipids, which according to the proposed model should fall within a physiological and metabolic health domain. The results of our study are thus counterintuitive as they suggest that these parameters are not associated with each other in this data set, which would be an unusual finding indeed.

Due to the lack of model fit, we used EFA to explore distinct health domains. Here, we observed that the TMS dataset was not factorable, indicating that there is no common variance among the different variables in the dataset. As a result, we removed the offending variables from the data set and reperformed the analysis, after verifying the reduced data set was factorable. The results of the EFA showed a two-domain model, distinguishing a *psycho-physical* domain, containing cognitive, physiological, social and physical performance measurements, and a domain containing measures of psychological resilience (refer to table 2). Not only does the result from our EFA differ from the five-domain model proposed by Lara *et al.,*[5] it also departs from our EFA conducted within the InChianti study. Here we show a clear four domain model, including neuro-sensory function, muscle function, cardio-metabolic function and adiposity.[20] Discrepancies between the different studies may be due to the cultural or behavioural differences among the participants of the different studies as well as to differences in how variables were assessed, especially in the context of self-rated items, analysed and age range

of the participants. Specifically, the InChianti study included a broad range of individuals whereas the TMS study an oversampling of T2DM. Interestingly, the original study regarding the HAP does not suggest any specific population, but merely produced a panel of measurements categorised into meaningful domains.

Although, some common themes (i.e. that health parameters can be grouped into meaningful domains) were observed between these studies and those from literature,[21] [20] the large discrepancies between the studies helps to demonstrate the complexity of defining and operationalising healthy aging.[22]

In a recent publication, operationalised models of successful ageing were reviewed. The authors reported that the models although multidimensional in nature, significantly vary in what they include. Of the 50 included studies, 19 models used Rowe and Kahn's standards of healthy ageing, two used the WHO's active ageing model and one study used Kuh's model of healthy biological ageing.[23] Moreover studies tend to use short surveys such as the SF-12 or SF-36 to measure healthy ageing instead of distinguishing separate health domains.[23] Although in this study we did not attempt to predict successful ageing, due to the cross-sectional nature of our data set, we aimed to test the possibility of creating a metric from a proposed panel. If we compare the HAP directly to another successful ageing, such as the Successful Aging Index (SAI), a four dimensional model containing (1) physical and functional health (PF), (2) psychological well-being and cognition (PC), (3) social engagement and family support (SF), (4) economic resources and financial security,[24] the models are similar in how aspects of health are grouped together. However, the HAP does not include measures of economic resources.[5]

Moreover, whereas the HAP[5] suggests a panel of metrics to measure these aspects of health, the SAI used a binary approach where participants could only score zero or one for each trait.[24] Interestingly in this study only 5.8% of participants attained SA in all four dimensions.[24]

Limitations

This study has a number of limitations. Since the data set had a significant number of missing values, a large number of participants (approximately 50% of the original data set) were lost during data analysis as CFA does not cope with any missing data. Similarly, due to the size of the complete data set, we were not able to replicate our analysis or results within a test data set. Moreover, we used a non-traditional analytical approach. Instead of performing an EFA followed by a CFA, our hypothesis was based on a proposed domain structure instead of exploring the data for its structure and then testing its robustness.

Conclusion

Our study shows that the previously published theoretical HAP model is not represented in the TMS cohort, which exemplifies the difficulty and complexities of defining, operationalising and measuring ageing success. Therefore, new studies should focus on developing an operationalizable metric of successful ageing which can then be tested within existing, and new cohort studies.

Declarations

Acknowledgements

None.

Conflicts of interest

None.

Funding
S.M. was funded by the Maastricht University Interfaculty Program ‘Eat well’.

References

1. Ory MG, Smith ML. Research, practice, and policy perspectives on evidence-based programing for older adults. *Frontiers in public health* 2015;3:136-36. doi: 10.3389/fpubh.2015.00136
2. Mount S, Lara J, Schols AM, et al. Towards a multidimensional healthy ageing phenotype. *Current opinion in clinical nutrition and metabolic care* 2016;19(6):418-26.
3. Ory MG, Smith ML. What If Healthy Aging Is the 'New Normal'? *International journal of environmental research and public health* 2017;14(11):1389. doi: 10.3390/ijerph14111389
4. Organization WH. World report on ageing and health: World Health Organization 2015.
5. Lara J, Godfrey A, Evans E, et al. Towards measurement of the Healthy Ageing Phenotype in lifestyle-based intervention studies. *Maturitas* 2013;76(2):189-99.
6. Schram MT, Sep SJ, van der Kallen CJ, et al. The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European journal of epidemiology* 2014;29(6):439-51.
7. Milligan GW, Cooper MC. A study of standardization of variables in cluster analysis. *Journal of classification* 1988;5(2):181-204.
8. Doornik JA, Hansen H. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* 2008;70:927-39.
9. White H. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society* 1982:1-25.
10. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability; 1967. University of California Press.
11. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal* 1999;6(1):1-55.
12. Rosseel Y. Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software* 2012;48(2):1-36.
13. Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria2008.
14. Hauben M, Hung E, Hsieh W-Y. An exploratory factor analysis of the spontaneous reporting of severe cutaneous adverse reactions. *Therapeutic advances in drug safety* 2017;8(1):4-16.
15. Acar S, Savci S, Keskinoğlu P, et al. Tampa Scale of Kinesiophobia for Heart Turkish Version Study: cross-cultural adaptation, exploratory factor analysis, and reliability. *Journal of pain research* 2016;9:445.
16. Stevens J. Applied multivariate statistics for the social sciences. 4th Edition ed. Mahwah, NJ: Lawrence Erlbaum Associates. 2002.
17. Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation* 2005;10(7):1-9.
18. Yong AG, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology* 2013;9(2):79-94.
19. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958;23(3):187-200.
20. Mount S, Ferrucci L, Wesselius A, et al. Measuring successful aging: an exploratory factor analysis of the InCHIANTI Study into different health domains. *Aging (Albany NY)* 2019;11(10):3023.
21. Parslow RA, Lewis VJ, Nay R. Successful Aging: Development and Testing of a Multidimensional Model Using Data From a Large Sample of Older Australians. *Journal of the American Geriatrics Society* 2011;59(11):2077-83. doi: 10.1111/j.1532-5415.2011.03665.x
22. Geard D, Reaburn PR, Rebar AL, et al. Masters athletes: Exemplars of successful aging? *Journal of Aging and Physical Activity* 2017;25(3):490-500.

23. Lu W, Pikhart H, Sacker A. Domains and measurements of healthy aging in epidemiological studies: A review. *The Gerontologist* 2018
24. Cheung KS-L, Lau BH-P. Successful aging among Chinese near-centenarians and centenarians in Hong Kong: a multidimensional and interdisciplinary approach. *Aging & mental health* 2016;20(12):1314-26.
25. Brinkhues S, Dukers-Muijrers NH, Hoebe CJ, et al. Socially isolated individuals are more prone to have newly diagnosed and prevalent type 2 diabetes mellitus-the Maastricht study–. *BMC public health* 2017;17(1):955.
26. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research* 1975;12(3):189-98.
27. Luteijn F, Barelds DPH. GIT2: Groninger Intelligentie Test 2: Harcourt Test Publishers 2004.
28. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure.[Research Support. *Non-US Gov't* 2001
29. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 2006;166(10):1092-97.
30. Bosscher RJ, Smit JH, Kempen GI. Algemene competentieverwachtingen bij ouderen: Een onderzoek naar de psychometrische kenmerken van de Algemene Competentieschaal (ALCOS). *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden* 1997
31. Pearlin LI, Schooler C. The structure of coping. *Journal of health and social behavior* 1978:2-21.
32. Bosma H, Van Jaarsveld C, Tuinstra J, et al. Low control beliefs, classical coronary risk factors, and socio-economic differences in heart disease in older persons. *Social science & medicine* 2005;60(4):737-45.
33. Klabbers G, Bosma H, van den Akker M, et al. Cognitive hostility predicts all-cause mortality irrespective of behavioural risk at late middle and older age. *The European Journal of Public Health* 2012;23(4):701-05.
34. Meesters C, Muris P, Bosma H, et al. Psychometric evaluation of the Dutch version of the Aggression Questionnaire. *Behaviour research and therapy* 1996;34(10):839-43.
35. Onete V, Henry RM, Sep SJ, et al. Arterial stiffness is associated with depression in middle-aged men—the Maastricht Study. *Journal of psychiatry & neuroscience: JPN* 2018;43(2):111.

Supplementary materials

Calculated variables

Social network variables: The collection of social network data has been described in detail elsewhere.[25] In brief, social networks were collected through a name generator questionnaire. The name generator first required the respondent to identify actual persons, and secondly asked several questions about these identified individuals are asked (i.e. sex, age, type of relationship, geographic distance, and the number of members who provided informational, practical or emotional support). For the present study the following social network variables were included;

1. Emotional support: The sum of two items "the number of network members that provide the opportunity to discuss important matters" and "the number of network members that provide emotional support when participants were feeling unwell".

2. Practical support: The sum of two items "the number of network members that help with small and larger jobs around the house" and "the number of network members that provide practical help when participants were sick"

3. Contact frequency (defined as an interaction between persons): Total contacts per half year were computed. For this, we first identified the highest contact frequency (e.g., daily contact) for every network member, as an indicator of the actual contact frequency. Secondly, we recoded the answer categories of the questionnaire to an estimated number of contacts per half year. For example, "bi-annually" was assumed to comprise one contact, "quarterly" two contacts, "monthly" 6 contacts and "daily or weekly" 48 contacts. Thirdly, we computed the sum of all contacts per half year as the total contact frequency, which was then divided by the total contact frequency by 10 for of the ease of interpretation.

MMSE score: Cognitive performance was assessed using the Minimal Mental State Examination.[26]

GIT score: "Groninger intelligentietest" word list test was used to assess verbal intelligence. [27]

Processing speed: Derived from the following (sub)tests: Stroop Colour Word Test Part I and II, Concept Shifting Test Part A and B (both adjusted for Part 0), and the Letter-Digit Substitution Test if all items had been completed and were otherwise coded as missing.

Executive attention: Derived from the following (sub)-tests: Stroop Colour Word Test Part III (interference score) and Concept Shifting Test Part C (inhibition score) if all items have been completed, otherwise they were coded as missing.

WLTR score: 15-word list delayed recall score.[6]

WLT score: 15-word list learning score.[6]

STRINT: Stroop test interference time (sec).

PHQ-9: Patient Health Questionnaire (PHQ-9)- nine item questionnaire was used to assess depressive symptoms.[28]

GAD-7: Generalised Anxiety Disorder (GAD-7) questionnaire used to assess symptoms of anxiety.[29]

GSES-effort: The General self-efficacy questionnaire sub-scale score calculation for "effort" as described in Bosscher et. al. 1997,[30] was calculated using items two, five, nine, twelve, fifteen, and sixteen if all items were answered by the participant. If any value was missing the calculated variable was not calculated and coded as missing.

GSES-Persistence: The General self-efficacy questionnaire sub-scale score calculation for "persistence" as described in Bosscher et. al. 1997,[30] was calculated using items one, three, four, eight, ten and thirteen if all items were answered by the participant. Of note, items one, three, four, eight, ten and thirteen were reverse coded. If any value was missing the calculated variable was not calculated and was coded as missing.

GSES initiative: The General self-efficacy questionnaire sub-scale score calculation for "Initiative" as described in Bosscher et. al. 1997, was calculated using items six, seven, fourteen and eleven, if all items were answered by the participant.[30] If any value was missing the calculated variable was not calculated and was coded as missing.

Personal mastery (mean): The personal mastery score was using a Dutch version of Pearlin and Schooler's Mastery scale[31], this is a seven item scale in which scores can vary from seven (low personal mastery) to 35 (high personal mastery).[32] This score was calculated as the mean of the responses so long as at least half the items on the questionnaire were answered.

Aggression score: The adult aggression questionnaire computes a score for aggression if all items have been answered.[33] Items are answered on a five point Likert scale, ranging from one ('totally disagree') to five ('totally agree') and summed, with higher scores indicating more cognitive hostility, anger and aggression, respectively. If any value was missing the calculated variable was not calculated and was coded as missing.

Of note, item one was reverse coded [34], and the original questionnaire also has a physical aggression component, but was not included in this study.

Physiological variables

Anthropomorphic measurements: Body mass index was determined from weight and height measurements; hip and waist measurements were also taken [6].

Blood pressure: Systolic and diastolic blood pressure were calculated as the average of at least three blood pressure readings (Omron 705IT) performed after a minimum of 10 minutes of seated rest.[35]

Blood lipids: Serum concentrations of total cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides were measured using an automatic analyser (Beckman Synchron LX20, Beckman Coulter Inc.).[35]

FEV1 percent predicted: Lung function, forced expiratory volume in one second, was assessed by spirometry (EasyOne™ spirometer, NDD Medical Technologies, Zurich, Switzerland).[6] Percent predicted was calculated using the Global Lung Function Initiative formula which accounts for age, height, gender and ethnicity.

Glucose: Fasting glucose concentrations were analysed using a standard enzymatic hexokinase reference method.[35]

HbA1c: HbA1c was collected as described elsewhere.[6]

Pulse pressure: Was calculated as the difference in systolic blood pressure and diastolic blood pressure.

Chapter 4

Measuring successful aging: An exploratory factor analysis of the InCHIANTI Study into different health domains

Authors:
Sarah Mount[1], Luigi Ferrucci[2], Anke Wesselius[3], Maurice P Zeegers[3], Annemie MWJ Schols[1]

[1]Department of Respiratory Medicine, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, the Netherlands

[2]Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, USA, Baltimore, MD 21224, USA

[3]Department of Complex Genetics, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, the Netherlands.

Abstract

Advocating continued health into old age, so called successful ageing, is a growing public health goal. However, the development of tools to measure ageing is limited by the lack of appropriate outcome measures, and operational definitions of successful ageing. Using exploratory factor analysis, we attempted to identify distinguishable health domains with representative variables of physical function, cognitive status, social interactions, psychological status, blood biomarkers, disease history, and socioeconomic status from the InCHIANTI study. We then used logistic and mixed effect regression models to determine whether the resulting domains predicted outcomes of successful ageing over a nine-year follow-up. A four-domain health model was identified: neuro-sensory function, muscle function, cardio-metabolic function and adiposity. After adjustment for age and gender, all domains contributed to the prediction of walking speed (R^2=0.73). Only the muscle function domain predicted dependency (R^2=0.50). None of the domains were a strong, significant predictor of self-rated health (R^2=0.18) and emotional vitality (R^2=0.23). Cross-sectional findings were essentially replicated in the longitudinal analysis extended to nine-year follow-up. Our results suggest a multi-domain health model can predict objective but not subjective measures of successful ageing.

Introduction

The number of old and very old adults (aged 65 and over, and 80 and over respectively) is rapidly rising in all European countries, and represents a progressively growing percentage of the general population.[1] At the same time, the proportion of working aged individuals is declining.[2] These changes in the population pyramid, as well as increasing life expectancy, is challenging the stability of health and social care systems.[3] Therefore, advocating strategies that promote health into old age and maximise *successful ageing* is a growing public health goal.

Biological ageing varies markedly between individuals,[4] and this disparity between individuals only grows with age.[5] Although partially genetically determined, 75% of human longevity is believed to be determined by modifiable factors including diet, lifestyle and socioeconomic status.[4] In order to understand whether any intervention aimed at promoting healthy ageing is effective, a benchmark for the assessment of healthy ageing is needed. Therefore, the development of tools to measure successful ageing, and to timely identify the early stages of health impairment, has become a research priority.[6] Developing such tools however, is a challenge, as ageing is a complex process and it is unlikely that a single measure will be able to track the ageing trajectory, particularly early in life, when disease symptoms and functional limitations are still rare.[6] Furthermore, testing the validity of tools to measure healthy ageing is complicated due to the lack of an agreed upon definition of healthy ageing,[6] as well as discrepancies in the terminology describing this concept.

As of 2010, 29 different definitions of successful ageing have appeared in the literature;[7] Michel & Sadana summarised the recent conceptualisations of ageing,[8] and more recently, a

citation network analysis identified 1146 publications related to successful ageing.[9] Despite extensive discussion and substantial analytical work done on defining successful ageing, many authors agree that an objective, robust measure of healthy ageing is difficult to develop. This is because at the individual level, successful ageing depends heavily on perspective of the observer, and different individuals may value different aspects of their life. [9] Subjective definitions tend to include themes such as the attainment or maintenance of goals, positive attitudes, attainment of social milestones and connectedness, whereas objective measures emphasise lack of disease and preserved functional status.[7] Yet, until a reliable measure of biological ageing can be developed, measures based on the aggregation of phenotypes, functional status, as well as subjective goals remain the best choice.

Previously defined key aspects of healthy ageing include physical and mental health as represented by walking speed,[10] dependency risk,[11] [12] emotional vitality [13] and self-rated health.[14] [15] Walking speed is a complex movement which integrates circulatory, respiratory, skeletal, muscular and nervous systems;[16] in older persons, it is a key indicator of physical health and a strong predictor of all-cause, cardiovascular, and other-cause mortality.[17-20] Dependency risk (deficits in the activities of daily living, ADL's), is strongly associated with severity of health status deterioration and a strong predictor of healthcare utilisation.[21] Emotional vitality is a subjective measure that summarises aspects of mental health, mood, psychological resilience and personal mastery,[13] all of which reflect the ability to adapt to changing personal circumstances,[22] physical changes,[13] and might contribute to the ability to find continued meaning in life. Lastly, self-rated health is a measure used widely in public health as an all-encompassing measure of health status,[23] thought to reflect brain-body communication.[24]

Due to the multidimensional nature of ageing, and age-related pathologies, assessing healthy ageing by combining information across many different measurements makes sense. Of the many proposed measurement tools, that at least in theory, assess health and wellbeing in older persons, most focus on disease and disability, which are only partial component of the multifaceted readouts of healthy ageing.[25] The problem with these tools is that they can easily distinguish the least healthy but not the healthiest individuals in the population.[26] This is because diseases and disabilities only become manifest when compensatory strategies are exhausted. Lara *et. al.* proposed that this problem could be addressed by operationally defining the Healthy Ageing Phenotype (HAP),[25] a panel of measures which change with age, are susceptible to lifestyle interventions and that can be classified in few meaningful domains. Clustering variables into domains may not only ease the interpretation of complex health data, it can also provide some clues of the underlying mechanism that affect the "healthy" condition. Furthermore, if the domains are identified using empirical methods, such as exploratory factory analysis, sub-scores can be developed that can capture changes in health status. Almost surprisingly, the development and use of such empirical methods have been limited to only a few studies.[27 28]

Using data from the InCHIANTI study, we aimed to identify distinguishable variable clusters (hereafter referred to as domains) that have face-validity for healthy ageing. We hypothesise that meaningful domains can be derived from the data, and that they are predictive of key aspects of successful ageing (Figure 1).

Methods

Study Design and participants

The data analysis was performed in the InCHIANTI database. InCHIANTI is a cohort survey that was designed and conducted to study risk factors and mechanisms of mobility loss in late life.[29] The initial data collection for this study began in September 1998 and the first phase was completed in 2000. Data collection continued thereafter every three years. For this study we used data collected at baseline (1998-2000) and in the three (2001 -2003), six (2004 -2006) and nine-year (2007 -2008) follow-up, which was concluded in 2009. Given the wide range of variables collected as well as the long follow-up, this cohort represents an extraordinary source for exploring factors associated with successful ageing. A detailed description of the InCHIANTI cohort study can be found elsewhere.[29] In short, 1453 adults, aged 20 and over were randomly selected from the population registries in two towns in the Chianti countryside of Tuscany, Italy Greve in Chianti and Bagno a Ripoli, which represented 94% of the eligible population.[30] The study was approved by the Italian National Institute of Research and Care of Aging ethical committee and complies with the Declaration of Helsinki.[30] The InCHIANTI study collected data on physical function, cognitive function, social status, dietary habits, psychological status, laboratory parameters, disease history, family history and socioeconomic status.[29] All analyse was performed in Stata 14.2.[31]

Variable selection

To select a putative list of variables that, at least in principle, could be potentially included as healthy ageing indicators in variable clusters, we first identified variables from the InCHIANTI dataset that had been previously associated with ageing and functional decline and had been included in other models of ageing and/ or allostatic load, including but not limited to the HAP. Part of this search entailed examining the models which were recently included in a review by

Mount *et. al*.[6] as well as looking at more recently developed models.[28 32-35] The final selection (table 1) was based on expert opinion by the research team. Correlation analysis was performed to remove redundant variables.

A total number of 1453 of subjects were included in our analysis. Of these, 44% were male and 66% were female. Females were on average 69 years, and males 67 years at baseline and age ranged from 23 to 97 years and 21 to 102 years for males and females respectively. A complete data set was available for 506 observations, which were subsequently included in the EFA, in order to avoid techniques such as multiple imputation.

Additionally, multivariate regression analyses were performed to explore if the predictive value and individual contribution of the four health domains at baseline was similar after nine-year follow-up. These tables are presented in the addendum (tables S6-S9).

Exploratory factor analysis

As a first step, to investigate the factor (domain) structure of the InCHIANTI dataset, we performed an exploratory factor analysis (EFA). We started by investigating multivariate normality using the Doornik-Hansen test[36] and the distributions of selected variables were explored by using histograms. Factorability of the data set was examined using an anti-image correlation matrix.[37 38] Values of 0.3 or higher were considered appropriate to be included in the EFA. EFA was then carried out using the principle axis factoring method with standardised variables. Variables were standardised using the variable value divided by maximum minus the minimum value method.[39] If variables had a loading of (-) 0.3 or higher, on at least one factor they were retained in the model.[40 41] In the construction of these domains, the factor loading

was carefully examined in the case of cross loadings. When variables had similar loadings on two factors, the variable factor was determined by logical relationships. In the case of HOMA, although it had a lower loading onto cardio-metabolic function, it was assigned to this factor as a result of testing both factor constructions. When assigned to the muscle function domain, it strongly reduced the reliability of the factor (0.65 to 0.56), while its addition to cardio-metabolic function had limited impact.

The resulting latent factors from the EFA were retained in the model based on the results of Kaisers criteria (eigen value greater than 1), as well as visual inspection of a scree plot (the number of factors to be retained in the solution is the number of factors which come before the elbow or levelling off of the curve).[41] In order to determine if the latent factors were correlated both varimax and promax rotations were performed on the resulting factor structure.[41] Secondly, a correlation analysis was performed on the resulting factors to verify the existence of any correlations between factors. The presence of any correlations and or differences between rotation methods determines the appropriate rotation method. Factors which consisted of at least three variables were considered stable.[41] To determine the sampling adequacy of the dataset, the Kaiser-Meyer-Olkin (KMO) was applied, as a rule of thumb this value should be greater than 0.6.[42] As a last step in the factor analysis, factor scores were then calculated using the predict function, a regression method in Stata, which were then used in further analysis.

Reliability testing

Internal consistency and reliability were examined using Cronbach's alpha for each of the extracted latent factors. Values greater than 0.9 were considered excellent, 0.8-0.9 good, 0.7-0.8 acceptable, 0.6-0.7 adequate, 0.5-0.6 poor and less than 0.5 as unacceptable.

Factoring scoring

Factor scoring coefficients were derived from the discovered latent factors using a regression method. The weights of the individual variables were then multiplied by the standardised measurements of individual participants to determine individual variable scores. These scores were then added to give an overall score to each of the individual latent variables.

Prediction models

Multivariate (logistic) regression analyses were used to determine the predictive value of the discovered factors (i.e. neuro-sensory function, muscle function, adiposity and cardio-metabolic function) on the key aspects of healthy ageing; self-rated health, walking speed, emotional vitality, and dependency at baseline. Model fit was examined by using R^2 in linear regression models and McKelvey and Zavoina's R^2 in logistic models [43]. In the analysis age at baseline and gender used as covariates.

Mixed effect regression and mixed effect logistic models were used to predict the dependent variables self-rated health, walking speed, emotional vitality, and dependency at the nine-year follow-up. Models were adjusted for baseline age and gender. To do this we calculated factor scores for each of the time points. If a variable was not available at a specific follow-up point, it was substituted for the value at the most recent follow-up moment. Once the factor scores were calculated, as previously described, they were entered into the model as independent

variables. In addition, a new variable, time point, which identified the factor scores at each time point, was included in the model as an independent variable.

Walking speed

Walking speed (m/s) was based on a 400m walking test. If the participants were not able to complete the test, the estimated 400m walking speed (m/s) was used.

Dependency

Participants were considered having disability if they had any need for help in performing Activities of Daily Living (ADL's), reflecting the lack of ability to perform the eating, bathing, dressing, toileting, transferring and maintaining continence unaided.[44]

Self-rated health

Participants were considered as having poor self-rated health if they state that they health was very-poor, poor or fair and to have good self-rated health if they self-reported that they health was good and very good.

Emotional vitality

Emotional vitality scores were generated following the method described by Penninx et. al. 2000,[13] with one exception. We had no complete measure of anxiety and therefore substituted the anxiety sub-score with the item from the CES-D questionnaire "During the past week, I felt fearful." Participants were given a score of zero if they scored more than one on this question, indicating they felt fearful more than rarely in the past week. Participants were considered vital if they if they passed all items (i.e. a score of four) but were otherwise considered not-emotionally vital.

Results

Exploratory factor analysis

Table 1 shows the factors resulting from the oblique rotation (table 1). All factors were used in further analysis (refer to supplementary table S1 for the orthogonally rotated factor loadings). Using the Eigen value criteria (table 1) as well as visual inspection of the scree plot (supplementary figure S1) four factors were identified defined as neuro-sensory function, cardio-metabolic function, muscle function and adiposity and were retained in the model (table 1, and figure 2). The factor set was then subjected to the KMO test, which had an overall value of 0.8447, indicating sufficient sampling adequacy.

Reliability testing

Neuro-sensory function, cardio-metabolic function, muscle function and adiposity variable clusters had, respectively, Cronbach's alpha coefficients of 0.78, 0.74, 0.65 and 0.55. Each of these factors remained stable when retested to derive the alpha value.

Factor scoring

The regression derived variable weights are presented in supplementary tables S2-S5.

Prediction models

Baseline: All domains were significant independent determinants of *walking speed* with a R^2 value of 0.78 (MSE 0.15) (table 2). Next to male gender, only muscle function was predictive of *dependency* with a R^2 of 0.50 (table 3). Predictive value for baseline *self-rated health* and *emotional vitality* models was low (R^2= 0.23 and 0.17 respectively (tables 4-5)), but it is of note that muscle function, neuro-sensory function and cardio-metabolic function contributed significantly to *self-rated health,* while only cardio-metabolic function contributed significantly to *emotional vitality*. An overview of the results is presented in table 6.

Nine-year follow-up: Similar results were obtained at the nine-year follow-up, except for the finding that future *dependency* was not only predicted by muscle function, but by all four domains (supplementary table S9).

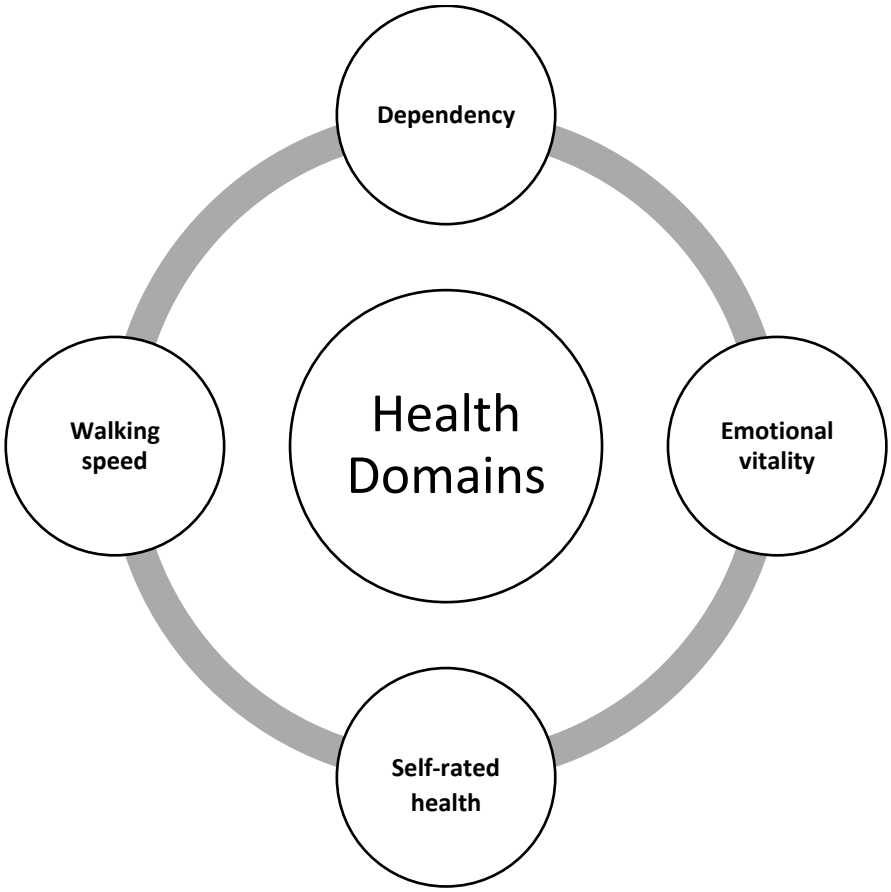


Figure 1: Theoretical model of successful ageing

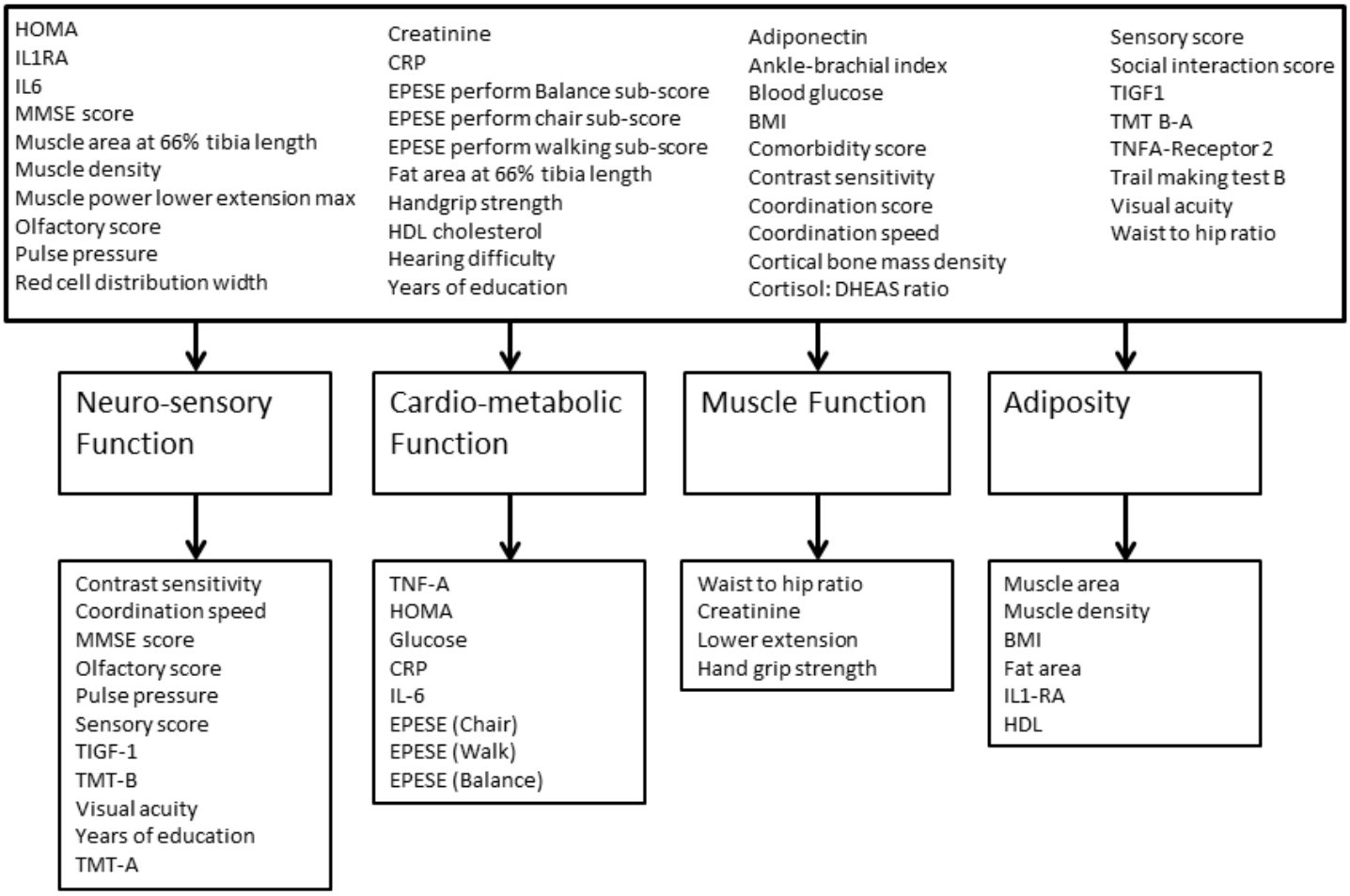


Figure 2: Extracted factor loading

Four factors (domains) named: Neuro-sensory function, Cardio-metabolic function, Muscle function and Adiposity.

Table 1

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Uniqueness
Adiponectin					0.8279
Fat area at 66% tibia length			0.4029	-0.8131	0.4144
Muscle area at 66% tibia length		0.3521	0.5086		0.6889
Muscle density	0.3936		-0.3447		0.6768
TNFA-Receptor 2		-0.3775			0.6192
HOMA		-0.3022	0.4763		0.6636
Blood glucose		-0.3355			0.7463
Creatinine				-0.5614	0.6918
Red cell distribution width					0.8993
Pulse Pressure	-0.4467				0.7458
Waist to hip ratio				0.4770	0.5632
EPESE perform walking sub-score		0.4925			0.6374
EPESE perform chair sub-score		0.4801			0.5802
EPESE perform Balance sub-score		0.5554			0.6092
Coordination score					0.9651
Coordination speed	0.6998				0.4397
Comorbidity score					0.7658
Muscle power lower extension max R side	0.5839			0.4267	0.3072
Trail making test B	-0.7686				0.4265
Years of education	0.7623				0.5005
Hearing difficulty					0.8857
IL6		-0.4676			0.7674
CRP		-0.3077			0.8350
IL1RA			0.4726		0.7427
Cortisol: DHEAS ratio					0.9549
Ankle-brachial index					0.9402
Cortical bone mass density					0.8337
HDL cholesterol			-0.4382		0.6506
TIGF1	0.4558				0.8022
Olfactory score	0.3925				0.7618
Sensory score	0.6311				0.5171
Social interaction score					0.8937
Handgrip strength	0.4989			0.4700	0.4041
BMI			0.8377	-0.3144	0.3811
Visual acuity	0.5421				0.6950
Contrast sensitivity	0.4651				0.6863
MMSE score	0.6832				0.5785

Oblique rotated factor loadings

Table 2

Walking speed	Coef.	Std. Err.	t	P>|t|	95% Conf. Interval	
Gender	-0.038	0.021	-1.850	0.065	-0.079	0.002
Age at baseline	-0.004	0.001	-6.090	0.000	-0.005	-0.003
Adiposity	-0.041	0.008	-5.110	0.000	-0.057	-0.026
Muscle function	0.140	0.014	10.050	0.000	0.112	0.167
Cardio-metabolic function	0.211	0.015	14.460	0.000	0.183	0.240
Neuro-sensory function	0.053	0.011	4.700	0.000	0.031	0.075
constant	1.536	0.043	36.120	0.000	1.453	1.620
Model information						
Observations	598	R^2 adj.	0.779			
p Model	0.000	Root MSE	0.147			
R^2	0.781	F (6, 691)	351.950			

Walking speed baseline model predictions

Table 3

Dependency	Coef.	Std. Err.	t	P>|t|	95% Conf. Interval	
Gender	7.284	7.340	1.970	0.049	1.010	52.503
Age at baseline	0.945	0.051	-1.050	0.293	0.851	1.050
Adiposity	1.431	0.511	1.000	0.316	0.711	2.881
Muscle function	0.117	0.118	-2.120	0.034	0.016	0.848
Cardio-metabolic function	0.521	0.194	-1.750	0.080	0.251	1.081
Neuro-sensory function	0.380	0.223	-1.650	0.100	0.120	1.202
constant	0.079	0.286	-0.700	0.484	0.000	97.444
Model information						
Observations	626	R^2	0.50			
p Model	0.000					
Log likelihood	35.05					

Dependency baseline model predictions

Table 4

Emotional vitality	Coef.	Std. Err.	t	P>|t|	95% Conf. Interval	
Gender	0.839	0.346	-0.430	0.671	0.374	1.883
Age at baseline	1.018	0.013	1.400	0.161	0.993	1.044
Adiposity	0.974	0.163	-0.160	0.876	0.702	1.352
Muscle function	1.107	0.295	0.380	0.701	0.658	1.865
Cardio-metabolic function	0.363	0.176	-2.090	0.037	0.140	0.940
Neuro-sensory function	0.830	0.202	-0.770	0.444	0.514	1.338
constant	3.294	2.756	1.420	0.154	0.639	1.697
Model information						
Observations	623	R^2	0.175			
p Model	0.000					
Log likelihood	32.610					

Emotional vitality baseline model predictions

Table 5

Self-rated health	Coef.	Std. Err.	t	P>|t|	95% Conf. Interval	
Gender	0.787	0.253	-0.740	0.456	0.420	1.477
Age at baseline	0.970	0.011	-2.770	0.006	0.949	0.991
Adiposity	1.128	0.140	0.980	0.330	0.885	1.438
Muscle function	0.549	0.129	-2.550	0.011	0.346	0.871
Cardio-metabolic function	0.481	0.108	-3.250	0.001	0.310	0.748
Neuro-sensory function	0.592	0.106	-2.930	0.003	0.418	0.841
constant	3.709	2.572	1.890	0.059	0.953	1.444
Model information						
Observations	623	R^2	0.225			
p Model	0.000					
Log likelihood	-328,918					

Self-rated health baseline model predictions

Table 6

	Walking Speed		Dependency		Self-rated health		Emotional vitality	
Domain	T0	T9	T0	T9	T0	T9	T0	T9
Gender		x	x			x		
Age at baseline	x	x		x	x	x		x
Adiposity	x	x		x		x		x
Muscle function	x	x	x	x	x	x		
Cardio-metabolic function	x			x	x		x	
Neuro-sensory function	x	x		x	x	x		
R^2	0.78	N/A	0.50	N/A	0.23	N/A	0.17	N/A

Relative latent factor contributions summary table. In this graphic baseline measurements are represented as T0, and the nine-year follow-up measurement as T9.

Discussion

There is a continued discussion in the literature as to what it means to age well, and terms vary from successful ageing, active ageing, positive ageing, productive ageing among others [45]. The aim of this study was to determine whether the phenotypic manifestation of ageing can be measured parsimoniously. Exploratory factor analysis using the InCHIANTI database lead to the discovery of four domains: Neuro-sensory function, cardio-metabolic function, muscle function and adiposity. Logical relationships were found between the variables making up the factors of cardio-metabolic function, muscle function, and adiposity. Neuro-sensory function

encompassed a compelling combination of measures of cognitive ability and sensory function, such as visual acuity and contrast sensitivity as well as other physiological factors. It was initially surprising that pulse pressure and insulin-like-growth-factor-1 (IGF-1) loaded on to this factor, but evidence from literature suggests strong and physiologically plausible relationships for this result. Previous longitudinal studies have found a relationship between higher pulse pressure and cognitive decline.[46] [47] Baseline pulse pressure, for instance, has been associated with poorer executive ability and lower total cerebral volume and greater temporal horn ventricular volume after five to seven years of follow-up.[48] This is supported by similar findings showing prospective declines in learning, nonverbal memory, working memory, and a cognitive screening measure among participants with increasing levels of pulse pressure.[47] Insulin-like-growth-factor-1 on the other hand, has been shown to decline with age and precede cognitive decline.[49] Furthermore IGF-1 has been shown to play a major role in growth, ageing, brain development and adult brain function,[49] and specific associations have been made with reductions in fluid intelligence,[50] and processing capacity.[51]

When these domains were then used to determine the key aspects of successful ageing, namely *walking speed*, *dependency risk*, *emotional vitality* and *self-rated health*, the directions of their contributions suggest that the domains are indeed useful. As summarized in table 6, both in the baseline and future walking speed model, high adiposity, and cardio-metabolic scores reduced walking speed, while high neuro-sensory and muscle function increased scores. In the baseline dependency model, only poor muscle function was predicted by dependency risk but, in the nine-year follow-up model, all domains became statistically relevant, with the strongest contributor being adiposity. These findings are in line with those of Diem *et. al.* who

found that maintained independence among the oldest age was related to mobility and cognitive function.[52]

The methods used here to develop a health score, and the outcomes we selected aid in avoiding the focus on 'average tendencies' within population subgroups, allow for heterogeneity and help shift the focus away from diseased and/or frail versus successful ager. Secondly, by carefully selecting the outcome variables we have avoided a focus on negative outcomes.[53] This also makes our model relevant to a wide range of the population by not limiting measures to those which are strictly related to frailty. In addition, by using only objective measurements, the influence of cultural differences may be reduced.[54] Lastly, by studying the ageing individual in this way allows us to consider that successful ageing may occur in the presence of (well managed) chronic disease[55] and recognises that ageing and its influence does not begin at any predefined cut off.

In general, our results lend support to the two schools of thought on successful ageing, specifically, the psychosocial school, which defines successful ageing as a mental state and the biomedical school which suggests successful ageing is avoidance of disease and disability.[56] Our model suggests physical aspects of ageing can be predicted well in contrast to emotional resiliency and one's health perspective.

Limitations

What we have shown here is that combining variables in the form of scores representing different systems can determine two aspects of what we consider successful ageing, namely walking speed and

dependency risk, both of which can importantly be influenced by lifestyle change. However, we should carefully consider the context in which this model was developed. Variables were selected from a pre-existing database, with preference for those which were available at multiple time points. On the other hand, these variables were also selected due to their consistent relationship with the ageing processes and were originally included in the database due to their possible relationship with disability.[29] Furthermore, we did not include early life factors, the impact of which is currently debated.[57] In addition, the factor analysis method should also be considered as the weightings of the specific variables and the composition of the factors may vary depending on the studied population. Our sample size also was limited because we chose to study a complete set of measurements. Lastly, we recognize that molecular metrics such as telomere length and methylation clock were not included as markers of biological ageing in the analysis. These measurements however are not normally done and cannot easily be added to typical blood panel chemistries. Furthermore, to date they are more theoretical instead of having practical use and for example, Haycock[58] elegantly demonstrated that telomere length remains controversial with respect to risk of cancer and non-neoplastic diseases.[58]

Conclusion

In a time of increasing longevity, reduced fertility rates, increased disease burden, as well as the availability of new and multiple alternative therapeutic opportunities, the ability to predict and measure the likelihood of an individual reaching old age, in a relatively good condition of health and wellbeing, is becoming progressively more important. From these considerations, our aim was to build a statistical that could help in building an objective operationalised definition of successful ageing based on data collected in large longitudinal study performed in a representative population. Our work clearly shows that combining complex measurements allow the prediction of future health outcomes within the domain of successful ageing. Our results show that parts of the ageing trajectory can be determined from a body systems

approach while others, specifically the components of healthy ageing that are more subjective, cannot. Future research could focus on improving this scale, or aspects of this scale within aspects of it such that we can predict the likelihood of maintained health, ability and emotional wellbeing into old age.

Declarations

Acknowledgements

None.

Conflicts of interest

None.

Funding

S.M. was funded by the Maastricht University Interfaculty Program 'Eat well'.

The InCHIANTI study baseline (1998–2000) was supported as a 'targeted project' (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the US National Institute on Ageing (contracts: 263 MD 916413 and 263 MD 821336); the InCHIANTI follow-up 1 (2001–2003) was funded by the US National Institute on Ageing (contracts: N.1-AG-1-1 and N.1-AG-1-2111); the InCHIANTI follow-ups 2 and 3 studies (2004–2010) were financed by the US National Institute on Ageing (contract: N01-AG-5-0002); this research was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Baltimore, Maryland.

References

1. Commission E. People in the EU – statistics on an ageing society Luxembourg: Eurostat; 2017 [updated 23 October 2017. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/People_in_the_EU_–_statistics_on_an_ageing_society accessed 23 October 2017.

2. Commission E. Population structure and ageing Luxembourg: Eurostat; 2017 [updated 20 July 2017. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing accessed 23 October 2017.

3. Berrío Valencia MI. Aging population: a challenge for public health. *Revista Colombiana de Anestesiología* 2012;40(3):192-94.

4. Collerton J, Barrass K, Bond J, et al. The Newcastle 85+ study: biological, clinical and psychosocial factors associated with healthy ageing: study protocol. *BMC geriatrics* 2007;7(1):14.

5. Martin P, Kelly N, Kahana B, et al. Defining successful aging: A tangible or elusive concept? *The Gerontologist* 2014;55(1):14-25.

6. Mount S, Lara J, Schols AM, et al. Towards a multidimensional healthy ageing phenotype. *Current opinion in clinical nutrition and metabolic care* 2016;19(6):418-26.

7. Jeste DV, Depp CA, Vahia IV. Successful cognitive and emotional aging. *World Psychiatry* 2010;9(2):78-84.

8. Michel J-P, Sadana R. "Healthy Aging" Concepts and Measures: Elsevier, 2017.

9. Kusumastuti S, Derks MG, Tellier S, et al. Successful ageing: A study of the literature using citation network analysis. *Maturitas* 2016;93:4-12.

10. Kuh D, Karunananthan S, Bergman H, et al. A life-course approach to healthy ageing: maintaining physical capability. *Proceedings of the Nutrition Society* 2014;73(2):237-48.

11. Lin P-S, Hsieh C-C, Cheng H-S, et al. Association between physical fitness and successful aging in Taiwanese older adults. *PloS one* 2016;11(3):e0150389.

12. Lowry KA, Vallejo AN, Studenski SA. Successful aging as a continuum of functional independence: lessons from physical disability models of aging. *Aging and disease* 2012;3(1):5.

13. Penninx BW, Guralnik JM, Bandeen-Roche K, et al. The protective effect of emotional vitality on adverse health outcomes in disabled older women. *Journal of the American Geriatrics Society* 2000;48(11):1359-66.

14. Cernin PA, Lysack C, Lichtenberg PA. A comparison of self-rated and objectively measured successful aging constructs in an urban sample of African American older adults. *Clinical gerontologist* 2011;34(2):89-102.

15. Whitley E, Popham F, Benzeval M. Comparison of the Rowe–Kahn Model of Successful Aging With Self-rated Health and Life Satisfaction: The West of Scotland Twenty-07 Prospective Cohort Study. *The Gerontologist* 2016;56(6):1082-92. doi: 10.1093/geront/gnv054

16. Studenski S, Perera S, Patel K, et al. Gait speed and survival in older adults. *Jama* 2011;305(1):50-58.

17. Nofuji Y, Shinkai S, Taniguchi Y, et al. Associations of walking speed, grip strength, and standing balance with total and cause-specific mortality in a general population of Japanese elders. *Journal of the American Medical Directors Association* 2016;17(2):184-e1.

18. Blain H, Carriere I, Sourial N, et al. Balance and walking speed predict subsequent 8-year mortality independently of current and intermediate events in well-functioning women aged 75 years and older. *The journal of nutrition, health & aging* 2010;14(7):595-600.

19. Dumurgier J, Elbaz A, Ducimetière P, et al. Slow walking speed and cardiovascular death in well functioning older adults: prospective cohort study. *Bmj* 2009;339:b4460.

20. Cooper R, Kuh D, Hardy R. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. *Bmj* 2010;341:c4467.

21. Ferrucci L, Guralnik JM, Pahor M, et al. Hospital diagnoses, Medicare charges, and nursing home admissions in the year when older persons become severely disabled. *Jama* 1997;277(9):728-34.

22. Barbic SP, Bartlett SJ, Mayo NE. Emotional vitality: concept of importance for rehabilitation. *Archives of physical medicine and rehabilitation* 2013;94(8):1547-54.

23. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social science & medicine* 2009;69(3):307-16.

24. Jarczok MN, Kleber ME, Koenig J, et al. Investigating the associations of self-rated health: heart rate variability is more strongly associated than inflammatory and other frequently used biomarkers in a cross sectional occupational sample. *PloS one* 2015;10(2):e0117196.

25. Lara J, Godfrey A, Evans E, et al. Towards measurement of the Healthy Ageing Phenotype in lifestyle-based intervention studies. *Maturitas* 2013;76(2):189-99.

26. Stute P, Bitterlich N, Bousquet J, et al. Measuring Active and Healthy Ageing: Applying a generic interdisciplinary assessment model incorporating ICF. *The journal of nutrition, health & aging* 2017;21(9):1002-09.

27. Caballero FF, Soulis G, Engchuan W, et al. Advanced analytical methodologies for measuring healthy ageing and its determinants, using factor analysis and machine learning techniques: the ATHLOS project. *Scientific reports* 2017;7:43955.

28. Araújo L, Ribeiro O, Teixeira L, et al. Predicting successful aging at one hundred years of age. *Research on aging* 2016;38(6):689-709.

29. Ferrucci L, Bandinelli S, Benvenuti E, et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *Journal of the American Geriatrics Society* 2000;48(12):1618-25.

30. Deshpande N, Metter EJ, Lauretani F, et al. Interpreting fear of falling in the elderly: what do we need to consider? *Journal of geriatric physical therapy (2001)* 2009;32(3):91.

31. StataCorp. 2015. Stata Statistical Software: Release 14. College Station TSL.

32. Carver LF, Buchanan D. Successful aging: considering non-biomedical constructs. *Clinical interventions in aging* 2016;11:1623.

33. Fugate Woods N, Rillamas-Sun E, Cochrane BB, et al. Aging Well: Observations From the Women's Health Initiative Study. *Journals of Gerontology-Series A Biological Sciences and Medical Sciences* 2016;71:S3-S12.

34. Goveas JS, Rapp SR, Hogan PE, et al. Predictors of optimal cognitive aging in 80+ women: the Women's Health Initiative Memory Study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 2016;71(Suppl_1):S62-S71.

35. Mejía ST, Ryan LH, Gonzalez R, et al. Successful aging as the intersection of individual resources, age, environment, and experiences of well-being in daily activities. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 2017;72(2):279-89.

36. Doornik JA, Hansen H. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* 2008;70(s1):927-39.

37. Hauben M, Hung E, Hsieh W-Y. An exploratory factor analysis of the spontaneous reporting of severe cutaneous adverse reactions. *Therapeutic advances in drug safety* 2017;8(1):4-16.

38. Acar S, Savci S, Keskinoğlu P, et al. Tampa Scale of Kinesiophobia for Heart Turkish Version Study: cross-cultural adaptation, exploratory factor analysis, and reliability. *Journal of pain research* 2016;9:445.

39. Milligan GW, Cooper MC. A study of standardization of variables in cluster analysis. *Journal of classification* 1988;5(2):181-204.

40. Stevens J. Applied multivariate statistics for the social sciences. 4th Edition ed. Mahwah, NJ: Lawrence Erlbaum Associates. 2002.

41. Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation* 2005;10(7):1-9.

42. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958;23(3):187-200.

43. Smith TJ, McKenna CM. A comparison of logistic regression pseudo R2 indices. *Multiple Linear Regression Viewpoints* 2013;39(2):17-26.

44. Katz S. Studies of illness in the aged. The index of ADL: a standardized measure of biologic and psychologic function. *JaMa* 1963;185:94-99.

45. Foster L, Walker A. Active and successful aging: A European policy perspective. *The Gerontologist* 2014;55(1):83-90.

46. Pérez-Zepeda MU, Belanger E, Zunzunegui MV, et al. Assessing the validity of self-rated health with the Short Physical Performance Battery: a cross-sectional analysis of the International Mobility in Aging Study. *PloS one* 2016;11(4):e0153855.

47. Waldstein SR, Rice SC, Thayer JF, et al. Pulse pressure and pulse wave velocity are related to cognitive decline in the Baltimore Longitudinal Study of Aging. *Hypertension* 2008;51(1):99-104.

48. Nation DA, Preis SR, Beiser A, et al. Pulse pressure is associated with early brain atrophy and cognitive decline: modifying effects of APOE4. *Alzheimer disease and associated disorders* 2016;30(3):210.

49. Frater J, Lie D, Bartlett P, et al. Insulin-like Growth Factor 1 (IGF-1) as a marker of cognitive decline in normal ageing: A review. *Ageing research reviews* 2018;42:14-27.

50. Aleman WRDV, Hans PF Koppeschaar, Mohamoud Osman-Dualeh, Harald JJ Verhaar, Monique M. Samson, Eduard Bol, Edward HF De Haan, André. Relationship between circulating levels of sex hormones and insulin-like growth factor-1 and fluid intelligence in older men. *Experimental aging research* 2001;27(3):283-91.

51. Aleman A, Verhaar HJ, de Haan EH, et al. Insulin-like growth factor-I and cognitive function in healthy older men. *The Journal of Clinical Endocrinology & Metabolism* 1999;84(2):471-75.

52. Diem SJ, Li-Yung L, Langsetmo L, et al. Effects of Mobility and Cognition on Maintenance of Independence and Survival Among Women in Late Life. *The Journals of Gerontology: Series A* 2017

53. Depp CA, Jeste DV. Definitions and predictors of successful aging: a comprehensive review of larger quantitative studies. *The American Journal of Geriatric Psychiatry* 2006;14(1):6-20.

54. Cosco TD, Prina AM, Perales J, et al. Operational definitions of successful aging: a systematic review. *International Psychogeriatrics* 2014;26(3):373-81.

55. Young Y, Frick KD, Phelan EA. Can successful aging and chronic illness coexist in the same individual? A multidimensional concept of successful aging. *Journal of the American Medical Directors Association* 2009;10(2):87-92.

56. Glass TA. Assessing the success of successful aging. *Annals of Internal Medicine* 2003;139(5_Part_1):382-83.

57. Wu S, Wang R, Zhao Y, et al. The relationship between self-rated health and objective health status: a population-based study. *BMC public health* 2013;13(1):320.

58. Telomeres Mendelian Randomization C, Haycock PC, Burgess S, et al. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA oncology* 2017;3(5):636-51. doi: 10.1001/jamaoncol.2016.5945

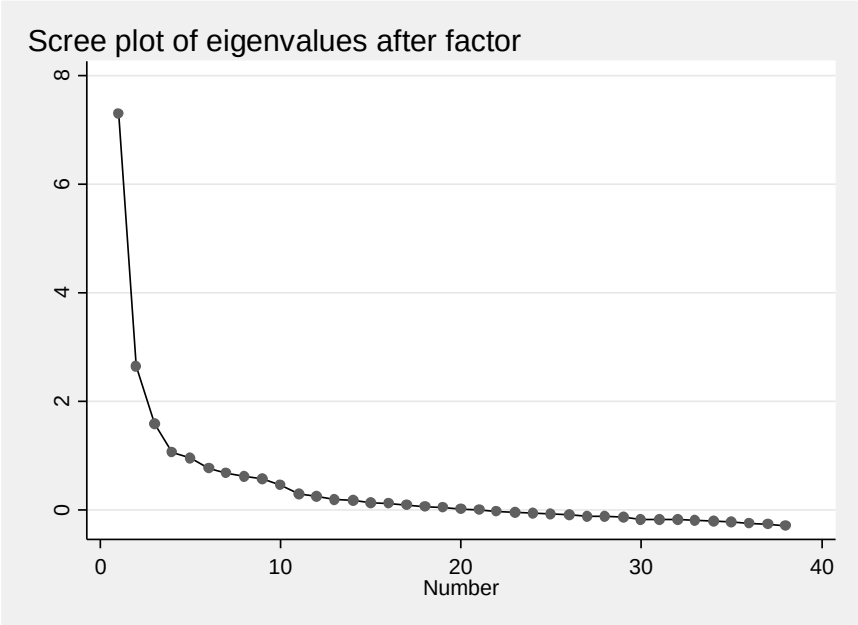


Figure S1: Scree plot

Scree plot demonstrating an elbowing of the curve at factor number three or four.

Table S1

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Uniqueness
Adiponectin		-0.3068			0.8279
Fat area at 66% tibia length			-0.7239		0.4144
Muscle area at 66% tibia length		0.4575			0.6889
Muscle density	0.4311	-0.3258			0.6768
TNFA-Receptor 2	-0.3755			-0.3793	0.6192
HOMA		0.5036			0.6636
Blood glucose		0.3533		-0.3170	0.7463
Creatinine			-0.5412		0.6918
Red cell distribution width	-0.3009				0.8993
Pulse Pressure	-0.4823				0.7458
Waist to hip ratio		0.3716	0.5119		0.5632
EPESE perform walking sub-score	0.3797			0.4470	0.6374
EPESE perform chair sub-score	0.4218			0.4348	0.5802
EPESE perform Balance sub-score	0.3433			0.4976	0.6092
Coordination score					0.9651
Coordination speed	0.7289				0.4397
Comorbidity score	-0.3855				0.7658
Muscle power lower extension max R side	0.6503		0.4477		0.3072
Trail making test B	-0.7464				0.4265
Years of education	0.6935				0.5005
Hearing difficulty					0.8857
IL6				-0.4217	0.7674
CRP					0.8350
IL1RA		0.4684			0.7427
Cortisol: DHEAS ratio					0.9549
Ankle-brachial index					0.9402
Cortical bone mass density	0.3487				0.8337
HDL cholesterol		-0.4921	-0.3040		0.6506
TIGF1	0.4341				0.8022
Olfactory score	0.4361				0.7618
Sensory score	0.6540				0.5171
Social interaction score					0.8937
Handgrip strength	0.5651		0.4749		0.4041
BMI		0.7464			0.3811
Visual acuity	0.5307				0.6950
Contrast sensitivity	0.5213				0.6863
MMSE score	0.6482				0.5785

Orthogonal (Varimax) rotated factor loadings

Table S2:

Neuro-sensory coefficients	
Pulse pressure	-0.074
Coordination score	0.208
Trail making test B	-0.238
Years of Education	0.147
TIGF1	0.064
Olfactory score	0.075
Sensory score	0.147
Visual acuity (near Snellen)	0.123
Contrast sensitivity	0.123
MMSE score	0.143

Regression generated scoring coefficients, Neuro-sensory

Table S3

Cardio-metabolic function coefficients	
TNFA- receptor 2	-0.103
HOMA index	-0.053
Blood glucose (mg/dl)	-0.051
z2PXSPSB	0.281
z2PXSPSC	0.234
z2PXSPSW	0.351
IL6	-0.119
CRP	-0.117

Regression generated scoring coefficients. Cardio-metabolic function

Table S4:

Muscle Function	
Serum creatinine (reciprocal)	0.084
Waist to hip ratio)	0.069
Muscle power lower ext. max R side(watts)	0.463
Handgrip strength	0.435

Regression generated scoring coefficients. Muscle function

Table S5

Adiposity coefficients	
Muscle area at 66% tibia length	0.129
Muscle density	-0.153
IL1RA	0.111
HDL	-0.111
BMI	0.564
Fat area at 66% tibia length	0.123

Regression generated scoring coefficients. Adiposity

Table S6

Walking speed	Coef.	Std. Err.	z	P>|z|	95% Conf. Interval	
Age	-0.012	0.000	-33.290	0.000	-0.012	-0.011
Gender	0.117	0.011	10.470	0.000	0.095	0.139
Neuro-sensory function	0.005	0.001	6.740	0.000	0.004	0.007
Cardio-metabolic function	-0.001	0.001	-1.310	0.189	-0.003	0.001
Muscle function	0.001	0.000	9.150	0.000	0.001	0.002
Adiposity	-0.050	0.013	-3.840	0.000	-0.076	-0.025
Constant	1.953	0.023	84.690	0.000	1.908	1.998
Random-effects Parameter var (Residual)						
Estimate	0.0325046		Wald chi2 (6)	1537.030		
Std. Err.	0.0013981		Log likelihood	318.082		
95% Conf. Interval	0.0298767	0.035364	Number of obs.	1081		
			Prob> chi2 >	0.000		

Mixed effect model for walking speed

Table S7

Emotional vitality	Coef.	Std. Err.	z	P>|z|	95% Conf. Interval	
Age	0.005	0.001	7.210	0.000	0.004	0.006
Gender	-0.042	0.022	-1.880	0.060	-0.085	0.002
Neuro-sensory function	-0.001	0.002	-0.580	0.564	-0.004	0.002
Cardio-metabolic function	-0.002	0.002	-0.850	0.398	-0.006	0.002
Muscle function	0.000	0.000	-0.800	0.425	-0.001	0.000
Adiposity	-0.066	0.024	-2.770	0.006	-0.113	-0.020
Constant	0.533	0.047	11.440	0.000	0.442	0.624
Random-effects Parameter var (Residual)						
Estimate	0.389		Wald chi2 (6)	64.060		
Std. Err.	0.008		Log likelihood	-608.974		
95% Conf. Interval	0.374	0.404	Number of obs.	1285		
			Prob> chi2 >	0.000		

Mixed effect model for emotional vitality

Table S8

Self-rated health	Coef.	Std. Err.	z	P>|z|	95% Conf.	Interval
Age	0.004	0.001	5.200	0.000	0.002	0.005
Gender	-0.103	0.024	-4.310	0.000	-0.149	-0.056
Neuro-sensory function	-0.004	0.002	-2.410	0.016	-0.007	-0.001
Cardio-metabolic function	-0.003	0.002	-1.330	0.184	-0.007	0.001
Muscle function	-0.001	0.000	-2.880	0.004	-0.002	0.000
Adiposity	0.069	0.026	2.700	0.007	0.019	0.120
Constant	0.095	0.050	1.900	0.058	-0.003	0.194
Random-effects Parameter var (Residual)						
Estimate	0.422		Wald chi2 (6)		727.050	
Std. Err.	0.008		Log likelihood		-727.050	
95% Conf. Interval	0.406	0.438	Number of obs.		1309	
			Prob> chi2 >		0.000	

Mixed effect model for self-rated health

Table S9

Dependency	Coef.	Std. Err.	z	P>|z|	95% Conf.	Interval
Age	0.002	0.000	5.630	0.000	0.001	0.002
Gender	-0.006	0.010	-0.560	0.577	-0.025	0.014
Neuro-sensory function	-0.002	0.001	-3.440	0.001	-0.004	-0.001
Cardio-metabolic function	-0.003	0.001	-2.620	0.009	-0.004	-0.001
Muscle function	-0.001	0.000	-3.360	0.001	-0.001	0.000
Adiposity	0.027	0.010	2.620	0.009	0.007	0.048
Constant	-0.050	0.021	-2.360	0.018	-0.092	-0.009
Random-effects Parameter var (sd Residual)						
Estimate	0.181		Wald chi2 (6)		77.650	
Std. Err.	0.003		Log likelihood		393.619	
95% Conf. Interval	0.175	0.188	Number of obs.		1361	
			Prob> chi2 >		0.000	

Mixed effect model for dependency

Chapter 5

Genetic, lifestyle, and environmental factors associated with development of chronic obstructive pulmonary disease: A systematic review and meta-analysis

Authors:

Sarah W Mount, MSc[1]*; Kelly F J Stewart, PhD[2,3]*; Anna-Roos Leerschool[2,3] MSc; Anke Wesselius, PhD[2]; Maurice P Zeegers, PhD[2,3]; Annemie W J Schols, PhD[1]

[1] Department of Respiratory Medicine, School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre Maastricht, Maastricht, The Netherlands

[2] Department of Complex Genetics and Epidemiology, School of Nutrition, and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre, Maastricht, The Netherlands

[3] CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre, Maastricht, The Netherlands

*Sarah Mount and Kelly Stewart contributed equally

Abstract

Although smoking is the key risk factor for Chronic Obstructive Pulmonary Disease (COPD), not all smokers develop it. Lifestyle, environmental and genetic factors may play a role in the aetiology. Our aim was to summarize the current state of knowledge, at the highest level of evidence, on factors associated with development of COPD. To achieve this goal, we reviewed relevant meta-analyses that we identified via MedLine, Web of Science and EMBASE. Meta-analyses were included if they reported genetic variants, lifestyle or environmental factors associated with development of COPD. For each risk factor - disease combination the most comprehensive odds ratio (OR) was determined by either calculation of a meta-OR through meta-meta-analysis, or by selecting the most comprehensive OR based on the most recent and/or complete meta-analysis. Of the articles we found, we selected 42 and 11 publications for genetic and lifestyle/environmental factors respectively, as most comprehensive or to be included in meta-meta-analysis. From these articles, 281 genetic variants were identified, 74% (n=208) showed a significant association with COPD with odds ratio's ranging from 0.17 to 3.33. Significant associations were found for seven of eight identified lifestyle/environmental factors, with odds ratio's ranging from 0.45 to 9.50.

Our report provides the first up-to-date and complete overview of genetic, lifestyle and environmental risk factors of COPD. This overview can serve as a valuable reference document for future researchers and lend itself for pathway analyses.

Introduction

Rational

Chronic obstructive pulmonary disease (COPD) is a preventable and treatable chronic respiratory disease characterized by persistent respiratory symptoms and airflow obstruction.[1] [2] Worldwide, it is estimated that 11.7% of people aged 30 and over suffer from the disease[3] and approximately 5% of all deaths is attributable to COPD; making it one of the leading causes of death.[4]

The most important and well-known risk factor for COPD is tobacco smoking,[5] [6] increasing the risk by over three-fold.[7] However, not all smokers develop COPD, only about 20% of smokers will ever develop the disease,[8] and furthermore not every COPD patient has a history of smoking behaviour.[6] Other typical non-genetic risk factors include the use of solid fuel for cooking and heating in poorly ventilated houses, as well as exposure to gases or fumes, mostly from occupational exposure.[9-11] In addition, a great number of genetic variants have been identified that were found to be associated with development of COPD.

With the large amount of evidence aggregated over the years regarding risk factors of COPD, a comprehensive summary at the highest level of evidence will be helpful to different end-users. To date, no such report has been published, incorporating genetic, lifestyle and environmental factors of COPD risk. Such an overview can help set up effective preventive strategies, screening practices, identifying knowledge gaps and supporting the work of future researchers.

Objective

It is our aim to give an overview of the current evidence of genetic, lifestyle and environmental factors associated with the development of COPD, as researched in meta-analyses.

Research Question

Which genetic mutations and lifestyle factors and or exposures significantly contribute to the development of COPD.

Methods

Prior to starting the research, a protocol was registered in PROSPERO (CRD42017051582).

Search strategy

Databases of MedLine, Web of Science, and EMBASE were searched from inception up to May 2018, with no restrictions to language or publication year. The following simplified search strategies were used: *(COPD OR emphysema OR chronic bronchitis) AND (Genome wide association study OR candidate gene studies OR polymorphism)* for genetic factors, and *(COPD OR emphysema OR chronic bronchitis) AND (diet OR dietary OR nutrition* OR exercise OR "physical activity" OR smoking OR tobacco OR environment* OR occupation* OR alcohol OR alcoholic) AND ("Meta-Analysis" OR "pooled analysis")* for lifestyle and environmental factors. The full search strategies can be found in appendix 1.

Eligibility criteria

Studies were eligible for inclusion when they reported genetic variants, lifestyle factors, or environmental factors associated with COPD risk in meta-analysis.

Study selection

After removal of duplicates, publications were screened by title and abstract. Second selection was based on full-text versions and done by SM and KS jointly. Disagreement was discussed until consensus was reached.

Data extraction

The following data was extracted: title, author, year, number of included primary studies, number of cases and controls, ethnicity, sex, relative risks per risk factor-disease combination (overall and subgroup), heterogeneity, and publication bias. Data was extracted by one researcher and checked for errors by a second researcher. Authors of publications were contacted for additional information in case of uncertainties.

Selection criteria for estimates

To be included, each meta-analytic estimate had to be based on a minimum of two primary studies. They were selected in the following order: (1) overall over sub-group estimates, (2) more adjusted over less adjusted estimates, (3) smoking adjustment over other adjustments, (4) incidence over prevalence, (5) random-effects model over fixed-effects model. Mortality as an outcome was only accepted if combined into one estimate with incidence and/or prevalence. Meta-analyses including both case-control and cohort studies were accepted.

Methodological quality

The Assessment of Multiple Systematic Reviews (AMSTAR), was used to appraise methodological quality of the individual meta-analyses.[12] Each included publication was rated separately by SM and KS and disagreements were solved through discussion. If consensus

could not be reached AW was consulted. The assessment was used for evaluative purposes only.

Data synthesis

Our goal was to include risk estimates for the most comprehensive study. If only one meta-analysis was available on a specific factor, this automatically was the most comprehensive study and these estimates were included directly. If multiple meta-analyses were available, the most comprehensive estimate was selected or calculated, in one of two ways: (1) Meta-analyses were combined in meta-meta-analysis to obtain an overall estimate based on the largest number of primary data. This was only considered appropriate if ≤50% of the studies of the smaller publication overlapped with those of the largest population. If this made meta-meta-analysis inappropriate, (2) the most comprehensive meta-analysis was included and secondly on the largest number of primary studies included. If this number was equal, the publication with the largest total number of participants was used. Meta-meta-analysis was performed using a random-effects model and heterogeneity was explored using the I^2-statistic.[13] All analyses were done using Stata 14.(Statacorp LP). The ORs from meta-meta-analysis will be referred to as a meta-OR (mOR).

Extracted estimates for genetic factors were categorized under their appropriate model: allelic (e.g. A vs. T), dominant (e.g. AA+AT vs. TT), recessive (e.g. AA vs. AT+TT), homozygote (e.g. AA vs. TT), heterozygote (e.g. AT vs. TT), over-dominant (e.g. AT vs. AA+TT), and 'other' (e.g. slow vs. normal enzyme activity).

Attributable risk percent and preventive fraction

To determine targets for preventive strategies, amongst other purposes, attributable risk per cent (AR%) was calculated for modifiable risk factors that were significantly associated with COPD. A combined AR% was calculated, combining lifestyle factors that are non-overlapping, assuming no interaction. This was done by first multiplying the individual RRs and then calculating the AR% as for individual factors. An AR% can be interpreted as the potentially preventable percentage of disease incidence among the exposed that are the result of the exposure, and therefore could be prevented if the exposure were eliminated.[14] In case of protective factors (RR<1.00), the RR was inversed to represent the risk associated with absence of this factor. ORs were used as an approximation of the RR if RRs were not available.

Results

Literature search

The literature search is depicted in Figures 1 and 2 for genetic and lifestyle/environmental factors respectively. A total of 60 and 15 publications were selected to be included in our study for genetic and lifestyle/environmental factors respectively. Of these, 42 and 11 publications were selected as most comprehensive for at least one risk factor or were included in meta-meta-analysis. Meta-meta-analysis was performed for two genetic and four lifestyle/environmental factors. Supplementary tables 1 and 2 present the selected publications, with relevant characteristics such as ethnicity, smoking behaviours and COPD diagnoses of included populations, and for which factors they have been selected as most comprehensive.

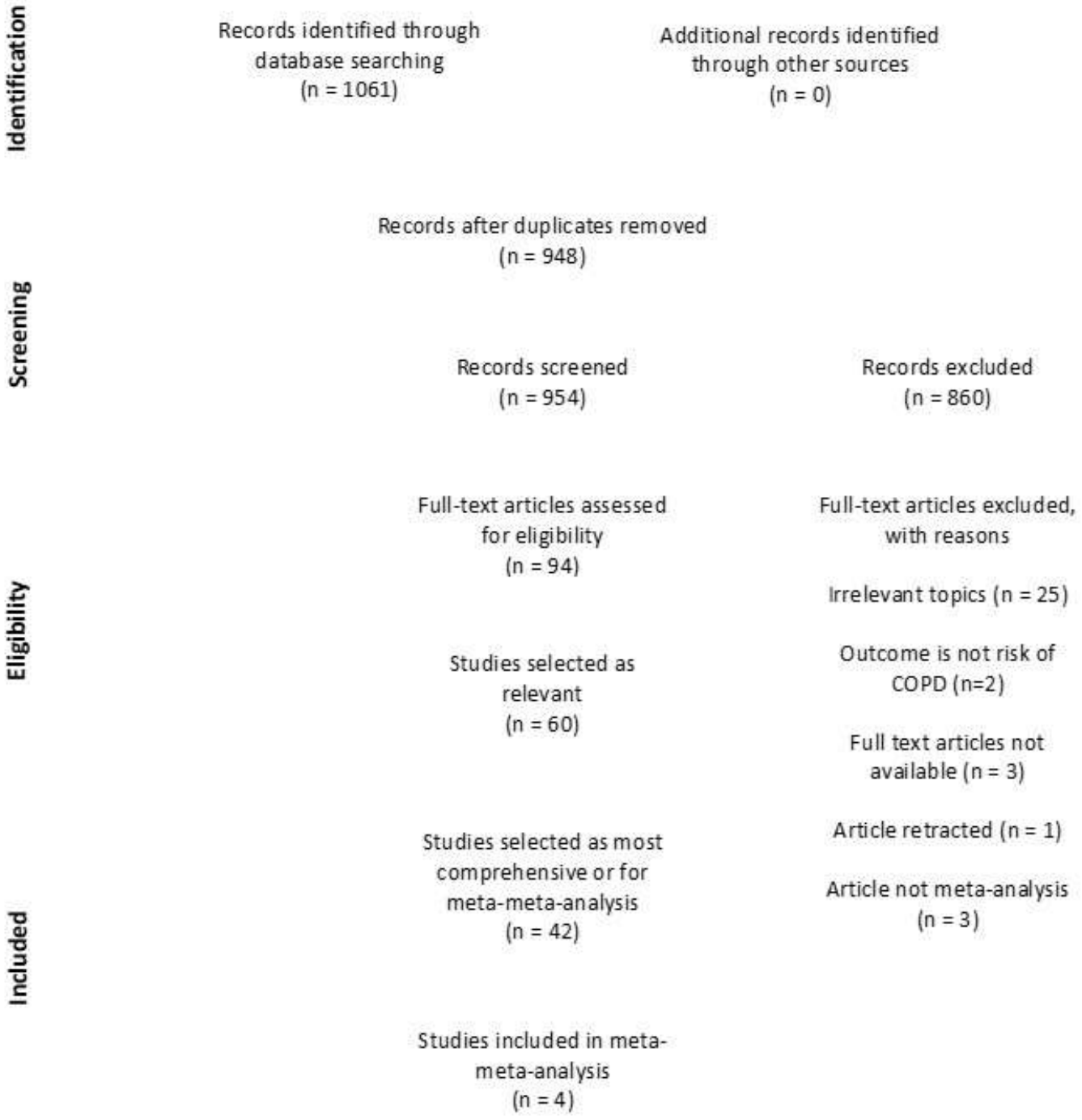


Figure 1: Flow diagram genetic factors

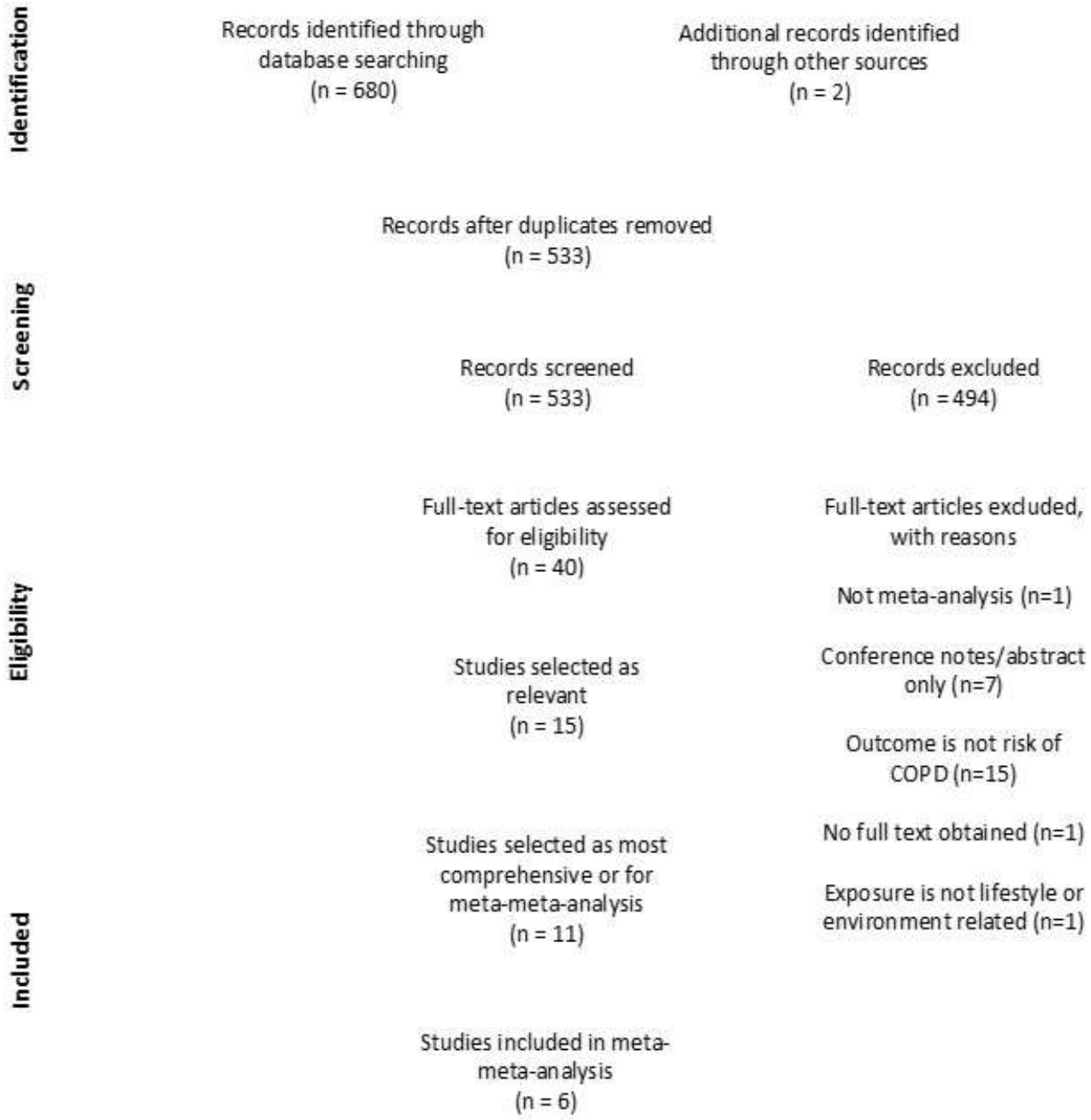


Figure 2: Flow diagram lifestyle factors

Reporting quality

Overall, AMSTAR ratings varied widely: 2/11 to 8/11 for meta-analyses and 0/4 to 3/4 for pooled analyses. This implies that the quality of the meta-analyses can be improved significantly. An overview of scores per item are visualized in supplementary figure 1 for genetic and lifestyle/environmental factors respectively.

Genetic factors

Of the 281 genetic variants identified, 74% (n=208) showed a significant association. Two meta-meta-analyses were performed, for CHRNA5 rs16969968 and IREB2 rs2568494. Both associations were statistically significant with meta-ORs of 1.30 (1.24-1.36) and 0.85 (0.76-0.93) respectively under the allelic model. Due to the large amount of information, only those variants with significant ORs ≤0.50 or ≥2.00 will be discussed and are additionally displayed in table 1 and figure 3. Variants with significant associations of 0.5>OR<2.0 and variants that were not significantly associated are displayed in supplementary tables 3 and 4 respectively.

Angiotensin I converting enzyme (ACE)

The insertion/deletion polymorphism increased the COPD risk in Asians, particularly in the homozygous and recessive models (OR=3.33, recessive model).[15] The risk was not increased in Europeans.

ADAM metallopeptidase domain 33 (ADAM33)

Many variants of the ADAM33 gene were associated with COPD, with particularly strong associations for rs2280090 (T2), rs2280091 (T1), and rs612709 (Q-1).[16] Rs2280090 (T2) increased the risk (e.g. OR=2.34, homozygous model) for Europeans, whereas the risk was

decreased for Asians (e.g. OR=0.17, homozygous model). Rs2280091 (T1) only showed an increased risk among Asians (OR=3.19, homozygous model). Rs612709 (Q-1) decreased the risk among both Europeans and Asians under all models (e.g. OR=0.51 and OR=0.28 respectively, homozygous model).

Cytochrome P450 family 1 subfamily A member 1 (CYP1A1)

The risk of COPD was particularly increased for rs1048943, but only when both ethnicities were combined (e.g. OR =2.75 and OR =3.23 for recessive and over-dominant models respectively).[17]

Epoxide hydrolase 1 (EPHX1)

Extremely slow activity EPHX1 encoded enzyme hydrolase 1 (determined by a combination of rs1051740 and rs2234922) resulted in an increased COPD risk in both the overall and Caucasian-specific estimates (OR=1.77 and OR=2.64 respectively), but not in Asians.[18]

Glutathione S-transferase mu 1 (GSTM1)

The null genotype of GSTM1 increased COPD risk for all ethnicities combined (OR=1.52, allelic model), but was particularly high for Africans (OR=2.42, allelic model) and Asians (OR=1.59, allelic model), in contrast to Caucasians (OR=1.26, allelic model) (Ding et al. 2018). Restricted to current and former smokers, estimates were similar (OR=1.51 and OR=1.59 for overall and Caucasians respectively, allelic model).[19] The risk was slightly higher for female than for male smokers (OR=2.74 and OR=2.04 respectively, allelic model).

Haeme oxygenase 1 (HMOX1)

Long (L) allele carriers of the HMOX1 gene had an increased COPD risk compared to short (S) and medium (M) allele carriers (e.g. OR=2.02 for people with at least one L-allele compared to non-L-allele carriers).[20]

Interleukin 6 (IL6)

The risk of COPD was decreased for rs1800796 (e.g. OR=0.20, homozygous model).[21]

Interleukin 13 (IL13)

Rs1800925 increased the COPD risk in the overall estimate (OR=1.82 and OR=2.02 heterozygous and homozygous models respectively) and specifically for Arabians (OR=2.94 and OR=3.05 heterozygous and homozygous models respectively).[22] Rs20541 decreased the risk in Caucasians (OR=0.38, dominant model) whereas it increased the risk in Asians (OR=1.30, allelic model).[23]

Interleukin 1 receptor antagonist (IL1RN)

The COPD risk was increased for rs2234663 in all ethnicities combined (OR=2.59 and OR=3.16, recessive and homozygous models respectively). A similar risk was seen for East-Asians (OR=3.20, homozygous model).[24]

Serpin family A member 1 (SERPINA1)

The Z allele of SERPINA1, compared to the M allele, increased the COPD risk by over twofold (OR=2.31) but this association disappeared after adjustment for smoking.[25] Compared to MM homozygotes, SZ heterozygotes had an increased risk of over threefold (OR=3.26).[26]

Surfactant protein A1/A2/B/D (SFTPA1/A2/B/D)

All variants of SFTPA1/A2/B/D combined led to an increased COPD risk for Asians (OR=1.97 and OR=2.26 for dominant and homozygous models respectively) but not Caucasians.[27]

Tumour necrosis factor (TNF)

Rs1800629 increased the COPD risk for Asians (e.g. OR=2.40 and OR=3.25 for allelic and homozygous models respectively) but not for non-Asians.[28] Restricted to smokers only, this association disappeared. Rs1800630 reduced the COPD risk in Asians only (e.g. OR=0.50, dominant model).[29]

Lifestyle and environmental factors

Publications were available on eight different lifestyle (Table 2) and environmental factors (Table 3). Only significant associations are discussed below and presented in figure 4.

Air quality

Different parameters of air quality in living environments was studied by Yang et al. (2017)[11] in a Chinese population. The risk of COPD was increased through living in a polluted area (OR=1.63) and poor housing ventilation (OR=3.99).

Active smoking

In total, five publications[7 30-33] examined COPD risks associated with active smoking. Forey et al. (2011)[7] was most comprehensive and was combined in meta-meta-analysis for several comparisons, with Jayes et al. (2016),[30] Wang et al. (2015)[32] and Kamal et al. (2015).[31]

Compared to never or non-smoking, risks for COPD were significantly increased for ever smoking (mOR=2.61), current smoking (mOR=3.51), and former smoking (mOR=2.44).

Risks differed by study design. The risk for ever-smoking was greatest in cohort studies, whereas for current and former smoking the risk was greatest in case-control studies. Stratification by sex showed that for all comparisons and outcomes risks were higher for males than females. Finally, stratification by country showed that, with some exceptions, risks were greatest in North-American studies.

Forey et al. (2011)[7] also studied other smoking-related parameters, such as intensity, duration, age at start smoking, and duration of quitting. In general, risks increased with increased intensity of smoking (all outcomes) and younger age at start smoking. Risks reduced with increased duration of quitting, although after 12 years of quitting the risk was still about a twofold increase compared to never smokers.

Passive smoking

The COPD risk for passive smoking was increased in females only (OR=2.17),[34] whereas the combined estimate was non-significant.[11 34]

Dietary patterns

High intake of an unhealthy or western diet (rich in red and processed meats, refined grains, sweets, desserts, and French fries) was associated with an increased risk (OR=2.12) whereas

high intake of a healthy or prudent diet (rich in vegetables, fruit, fish, and wholegrain) decreased the risk (OR=0.55; inverse association).[35]

Solid fuel smoke

In total, five publications[9 11 36-38] examined solid fuel smoke exposure related to COPD risk. Kurmi et al. (2010)[9] was most comprehensive for any solid fuel smoke and wood smoke. Sana et al. (2018)[38] was most comprehensive for biomass exposure in females and could be combined with Yang et al. (2017)[11] in MMA for the most comprehensive estimate for both males and females combined. All types of solid fuel smoke combined increased the risk of COPD with an OR = 2.80. Sub-estimates by type of fuel showed that particularly wood smoke increased the risk (OR = 4.29) and no significant association for biomass smoke exposure.

Occupational exposure to vapours, gases, dusts or fumes

In total, four publications[10 11 39 40] reported on exposure to vapours, gases, dusts or fumes in relation to COPD, with Sadhra et al. (2017)[10] being most comprehensive. This publication was combined with Yang et al. (2017)[11] for exposure to dusts. The risk of COPD was increased in the overall estimate (OR=1.22), and was higher among females than males (OR=1.78 and OR=1.32 respectively). Stratified by study design, the risk was lower for cohort studies (OR=1.11) than for case-control (OR=1.75) or cross-sectional studies (OR=1.21). Furthermore, when stratified by level of exposure, the risks were only significant for high exposures (OR=1.36), but not at low or medium level exposures. Finally, stratified by type of exposure, risks were significantly increased for COPD for vapours (OR=1.24), gases (OR=1.10), dusts (mOR=1.38), biological dusts (OR=1.33), mineral dusts (OR=1.07), and fumes (OR=1.16).

Waterpipe tobacco smoking

Waterpipe tobacco smoking increased the risk of COPD by over threefold (OR=3.18)'.[41]

Attributable risk percent

The AR% was calculated for each lifestyle and environmental/occupational factor (Tables 2 and 3). Results show that among all current smokers 72% of COPD cases could have been prevented, and among those smoking around 45 cigarettes per day this rose to 89%. In addition, 53% of COPD cases among those with poor (Western-style) diets could have prevented. These two factors (current smoking and poor diet) were combined to calculate a combined AR%, showing that 87% of cases could have been prevented among smokers with a poor diet. With regard to environmental factors it appears that 64% of cases who have been exposed to solid fuel smoke and 74% of cases living in housing with poor ventilation could have been prevented.

ACE
Ins/Del
Asian: 3.33
European: NS

ADAM33
Rs2280090
Asian: 0.17
European: 2.34

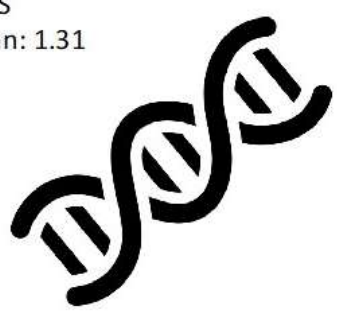
Rs2280091
Asian: 3.19
European: NS

Rs612709
Asian: 0.28
European: 0.51

CYP1A1
Rs1048943
Asian & Caucasian: 3.23

EPHX1
Extremely slow enzyme activity
Asian: NS
Caucasian: 2.64

Slow enzyme activity
Asian: NS
Caucasian: 1.31



GSTM1
Null/Wildtype
African: 2.42
Asian: 1.59
European: 1.26
Females: 2.74
Males: 2.04

HMOX1
L/S+M
Asian & Caucasian: 2.02

IL6
Rs1800796
Asian & Caucasian: 0.20

IL13
Rs1800925
Asian: NS
Caucasian: NS
Arabians: 3.05

IL1RN
Rs2234663
East-Asian, South-Asian & Arabian: 3.16

SERPINA1
Z/M
Caucasian: 2.31
Smoking-adjusted: NS
SZ/MM
Caucasian: 3.26

SFTPA1/A2/B/D
Wildtype/Mutant
Asian: 2.26
Caucasian: NS

TNF
Rs1800629
Asian: 3.25
Non-Asian: NS
Smokers-only: NS

Rs1800630
Asian: 0.50
Non-Asian: NS

The presented ORs/RRs may be for different genetic models. For detailed results please see table 1.

Figure 3: Summary figure genetic factors

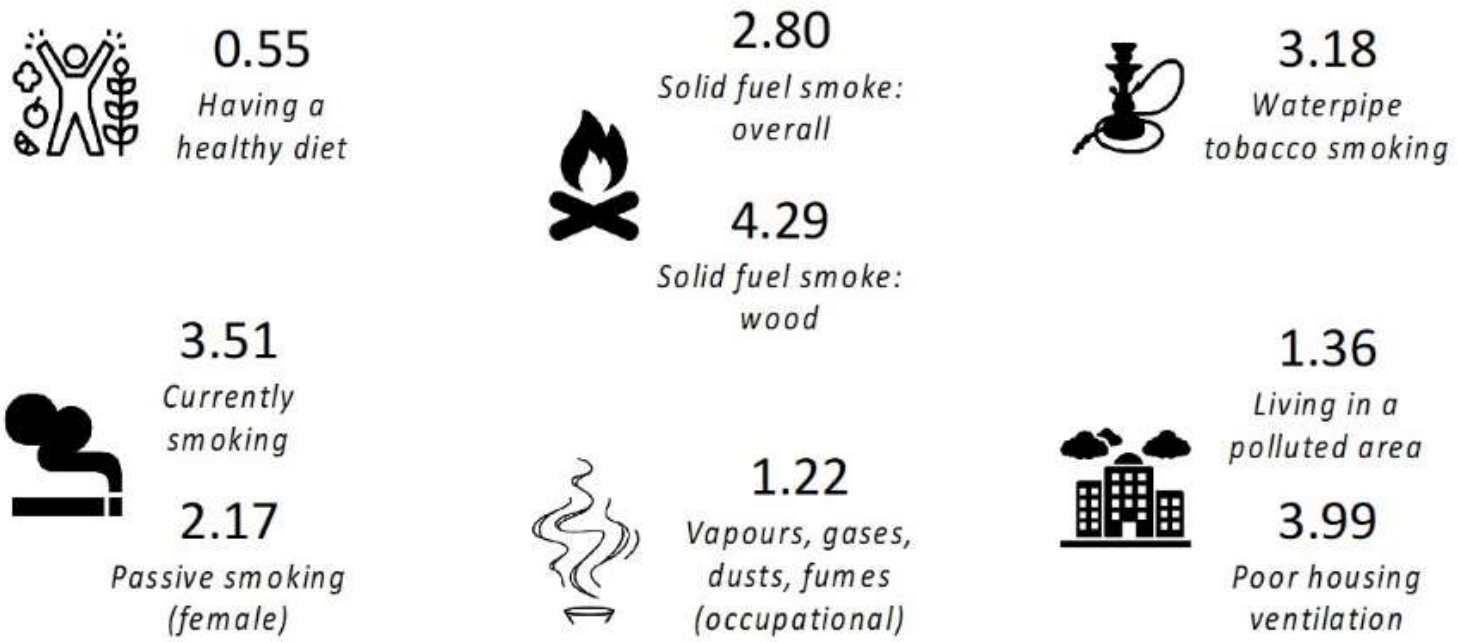


Figure 4: Summary figure lifestyle factors

Discussion

This review highlights that COPD is not a single-cause disease and multiple different factors should be considered to understand the origins of COPD.[42] To our knowledge, this is the first systematic review and meta-meta-analysis of genetic, lifestyle, and environmental factors associated with COPD risk. It provides a comprehensive, reliable, and up-to-date summary of current knowledge at the highest level of evidence. The identified risk factors, both genetic and non-genetic, can mostly be placed within several key pathways of COPD development: detoxification, addiction and chronic inflammation.

Firstly, exposure to toxic substances, such as those found in tobacco smoke or solid fuel smoke, are well-known risk factors of COPD. In fact, smoking is one of the riskiest behaviours with regard to COPD.[7 30-32] In order for our bodies to detoxify such substances, well-functioning enzymes are required. Deleterious mutations in genes involved in the detoxification pathways can therefore influence the risk of COPD. For example, EPHX1 is involved in the detoxification of polycyclic aromatic hydrocarbons (PAHs) amongst others. PAHs are present in cigarette smoke and (indoor) air pollution,[43] and have been found to be related to the development of COPD.[44 45] Extremely slow variants of the by EPHX1 encoded enzyme epoxide hydrolase 1 increased the COPD risk by 77%.[18] This suggests an interaction between environmental and genetic factors will be present, with individuals with slower enzyme activity being at an even greater increased risk if exposed to toxic substances compared to individuals with normal or even fast enzyme variants.

Secondly, variants in genes involved in addictive behaviours are also likely candidates to increase disease risk, as they may increase exposure to toxic substances such as from smoking. A great number of SNPs have been identified in the CHRNA3 and CHRNA5 genes.[46-48] encoding a nicotinic cholinergic receptor.[11] Variants in these genes may lead to an exaggerated pleasure in response to smoking and thereby increasing the chances of repeated exposure.[49] Similarly, the MMP genes mediate a variety of biological functions including a role in the body's reward system, aiding in potentiating the rewarding effects of drugs[50 51] as well as playing an inflammatory role.[52] Our results show a significant association for MMP9 rs17576 and rs3918242, but not for any variants in MMP1, MMP3 or MMP12.

Thirdly, chronic inflammation plays a critical role in the pathogenesis and progression of COPD.[53] Many genes dysregulated by exposure to cigarette smoke are those related to oxidative stress and central components of inflammatory signalling pathways.[54] One well-known COPD-causing gene is the SERPINA1 gene, encoding α1-antitrypsin (AAT). AAT protects the alveoli from neutrophil elastase and other proteases.[55] However, if this gene is mutated it may result in slow versions of AAT, leading to AAT-deficiency. As a result, individuals develop emphysema and early-onset or more severe forms of COPD.[56] The deficient S and Z alleles (slow and very slow enzyme activity respectively) of SERPINA1 increased the risk of COPD.[25 26] Interestingly, for several comparisons in this gene, the association disappeared after adjustment for smoking. This is contradictory to other research as well as information commonly communicated to patients.[56 57] A similar observation was seen for TNF. It could be explained as the result of upregulation of other genes involved in the same pathway as a result of the frequent presence of cigarette smoke.[58] However, it may also be a result of

methodological issues. Future research should focus on gene-environment interactions and pathways by which these effects occur. At the moment, evidence on gene-environment interactions is only slowly accumulating, likely as a result of a limitation due to methodological problems such as limited sample sizes, variant prevalence, and reproducibility of their findings.[59]

In contrast to pro-inflammatory genetic mutations, several nutrients may have anti-inflammatory properties. Literature suggests that antioxidants including certain omega-3 fatty acids, vitamins and minerals found in fruits and vegetables may mitigate the inflammatory effects of smoking. A recent review showed evidence for the positive influence of omega 3 fatty acids and vitamin E, which is at least partly attributable to their anti-inflammatory actions.[60] Intake of a healthy diet significantly reduced the risk of COPD with an OR of 0.55,[35] further substantiating this argument.

Further interactions exist with ethnicity or sex. For example, the rs2280090 (T2) of the ADAM33 gene showed an increased risk among Europeans but decreased risk for Asians. Furthermore, several variants in the FOXO1 gene showed a decreased risk among female smokers only.[46] An explanation for these ethnicity and sex-related differences may be due to different allele frequencies or exposure levels between populations or groups, (yet unknown) gene-environment interactions, linkage disequilibrium or remaining confounding factors.[16]

Interestingly, and unexpectedly considering the importance of active smoking as a risk factor for COPD, our MMA results showed that passive smoking was not associated with COPD. In addition, we noted little evidence on prenatal and childhood exposure to passive tobacco smoking in relation to COPD risk. Such a link is not unlikely given the results of previous research, which report poorer lung function and increased risk of chronic respiratory disease in offspring exposed to tobacco smoke prenatally or during childhood,[61-63] but studying this topic is methodologically and ethically difficult.

By calculation of an attributable risk percent, we were able to give an intuitive interpretation of how many cases could be prevented if it were possible to introduce or eliminate a particular behaviour in the entire population. For example, among individuals with a poor diet (referred to as a Western-style diet, low in fruit and vegetables, high in processed meats, sugars and saturated fats) 53% of COPD cases could have been prevented. Similarly, 72% of cases among smokers could have been prevented. Together with information on prevalence of these exposures, this can guide public health strategies.

Limitations

Our results are currently the most comprehensive available, yet they should be interpreted with caution. We urge the reader to critically appraise the results of each meta-analysis, when applying our results to their own purpose. Firstly, the quality of our results is dependent on the quality of the included meta-analyses. We judged their quality using the AMSTAR checklist, including all studies regardless of their score. These ratings showed that quality can be improved. For example, many studies did not do a comprehensive literature search, limiting to

only one database and not searching grey literature, which may have introduced publication bias.[64] Also, only a few studies provided a published 'a priori' design. Secondly, it is essential that the meta-analyses we included sufficiently deal with the quality of their included studies, by both evaluating it and taking the result into consideration when formulating conclusions. Thirdly, not all studies reported risks by smoking status and primary studies had different baseline smoking exposures. As smoking is such a major risk factor for COPD, this information may be critical when generalizing these results to the general population, for example in their use for gene-environment interaction analyses, pathway analyses, or the development of preventive strategies and genetic risk scores. Even more so, baseline smoking exposure will be particularly relevant when the gene is involved in smoking-related metabolism. Fourthly, we have not checked potential linkage disequilibrium between genetic variants found in different meta-analyses. If linkage disequilibrium is present, two significantly associated variants may actually be the effect of one single variant. Fifthly, as a result of the increased risk of a type II error in GWA studies, correction for multiple testing and therefore a very high statistical significance needs to be achieved for variants to be recognized as risk factors in GWA studies. However, as a result, some actual predictive variants may have been missed and it is possible that different variants may be identified in future studies.[42] Sixthly, COPD diagnosis methods varied greatly between the included studies and not all studies had spirometric confirmation of diagnosis. Also, different COPD phenotypes may have been included,[65] and risk factors may play different roles in each phenotype. Future research may shed light on the relative relevance of the different risk factors in different phenotypes. Finally, few meta-analyses could be combined in meta-meta-analysis due to great amount of overlap in primary studies between them, which would result in false precision and homogeneity if combined.[66] Even so, taken

together, our results still represent a valuable overview of the most comprehensive estimates currently available.

Conclusion

A large number of genetic, lifestyle, and environmental factors have been identified that each contribute to the development of COPD, either independently or through mutual interactions. Strongest genetic associations were found forADAM33 and the PI SZ genotype of SERPINA. Among the lifestyle factors, adopting a healthy diet, and avoiding active, passive and waterpipe smoking can result in great risk improvements. Based on our findings we suggest that future research should focus on gene-environment and gene-gene interactions.

The data collected in this overview lends itself to pathway analysis, which integrates interactions of genes and metabolites in specific biological processes.[67] Using our data in such context will provide new insight into the pathway of COPD development, the relationships with other chronic diseases and improve disease outcome prediction models.[68]

Declarations

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

All authors designed and approved of the research protocol. SM and KS did literature search, assessed eligibility of studies for inclusion, and extracted data. AW was consulted in case of

disagreement. AL additionally checked and reviewed the extracted data to solve any uncertainties in risk and reference alleles. KS did the statistical analysis in consultation with SM and AW. SM and KS wrote the first draft of the article. All authors critically revised the manuscript for intellectual content, finally approved of the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

This study was funded by the Maastricht University Interfaculty Program 'Eatwell'.

Acknowledgments

This study was funded by the Maastricht University Interfaculty Program 'Eatwell'. Images used to create figures 3 and 4 were obtained from the Noun Project and include the following: DNA by Delwar Hossain, Healthy lifestyle by Becris, Smoking by jungsang, Environmental pollution by ProSymbols, Campfire by Adrien Coquet, Shisha by bmijnlieff, and Smoking fumes by Olena Panasovska.

References

1. Guirguis-Blake JM, Senger CA, Webber EM, et al. Screening for chronic obstructive pulmonary disease: evidence report and systematic review for the US preventive services task force. *Jama* 2016;315(13):1378-93.
2. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management and prevention of chronic obstructive lung disease 2017 report. *Respirology* 2017;22(3):575-601.
3. Adeloye D, Chua S, Lee C, et al. Global and regional estimates of COPD prevalence: Systematic review and meta–analysis. *Journal of global health* 2015;5(2)
4. Alwan A. Global status report on noncommunicable diseases 2010: World Health Organization 2011.
5. Raherison C, Girodet P. Epidemiology of COPD. *European Respiratory Review* 2009;18(114):213-21.
6. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. *The Lancet* 2015;385(9971):899-909.
7. Forey BA, Thornton AJ, Lee PN. Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema. *Bmc Pulmonary Medicine* 2011;11 doi: 10.1186/1471-2466-11-36
8. Bosse Y. Updates on the COPD gene list. *International journal of chronic obstructive pulmonary disease* 2012;7:607-31. doi: 10.2147/copd.s35294 [published Online First: 2012/10/12]
9. Kurmi OP, Semple S, Simkhada P, et al. COPD and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis. *Thorax* 2010;65(3):221-28. doi: 10.1136/thx.2009.124644
10. Sadhra S, Kurmi OP, Sadhra SS, et al. Occupational COPD and job exposure matrices: a systematic review and meta-analysis. *International journal of chronic obstructive pulmonary disease* 2017;12:725-34. doi: 10.2147/copd.s125980
11. Yang Y, Mao J, Ye Z, et al. Risk factors of chronic obstructive pulmonary disease among adults in Chinese mainland: A systematic review and meta-analysis. *Respiratory Medicine* 2017;131:158-65.
12. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology* 2007;7(1):1.
13. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal* 2003;327(7414):557.
14. Bouter LM, Zielhuis GA, Zeegers MP. Textbook of Epidemiology. 1 ed: Bohn Stafleu van Loghum 2018.
15. Kang SW, Kim SK, Chung JH, et al. Genetic Polymorphism of Angiotensin-Converting Enzyme and Chronic Obstructive Pulmonary Disease Risk: An Updated Meta-Analysis. *BioMed Research International* 2016;2016 (no pagination)(7636123)
16. Zhou DC, Zhou CF, Toloo S, et al. Association of a disintegrin and metalloprotease 33 (ADAM33) gene polymorphisms with the risk of COPD: an updated meta-analysis of 2,644 cases and 4,804 controls. *Molecular biology reports* 2015;42(2):409-22. doi: 10.1007/s11033-014-3782-5 [published Online First: 2014/10/05]
17. Wang CD, Chen N, Huang L, et al. Impact of CYP1A1 Polymorphisms on Susceptibility to Chronic Obstructive Pulmonary Disease: A Meta-Analysis. *BioMed Research International* 2015;2015 (no pagination)(942958)

18. Li H, Fu WP, Hong ZH. Microsomal epoxide hydrolase gene polymorphisms and risk of chronic obstructive pulmonary disease: A comprehensive meta-analysis. *Oncology letters* 2013;5(3):1022-30. doi: 10.3892/ol.2012.1099 [published Online First: 2013/02/22]
19. Xue H, Su J, Sun K, et al. Glutathione S-transferase M1 and T1 gene polymorphism and COPD risk in smokers: an updated analysis. *Molecular biology reports* 2012;39(4):5033-42. doi: 10.1007/s11033-011-1300-6 [published Online First: 2011/12/14]
20. Zhou H, Ying X, Liu Y, et al. Genetic polymorphism of heme oxygenase 1 promoter in the occurrence and severity of chronic obstructive pulmonary disease: a meta-analysis. *Journal of cellular and molecular medicine* 2017;21(5):894-903.
21. Ji G-Y, Tang X-Y, Wu S-Q, et al. Association between the IL-6 rs1800796 (-572C/G) polymorphism and chronic obstructive pulmonary disease risk: a meta-analysis. *Int J Clin Exp Med* 2017;10(7):10375-81.
22. Chen L, Shen Y, Liu L, et al. Interleukin-13 -1112 C/T promoter polymorphism confers risk for COPD: a meta-analysis. *PloS one* 2013;8(7):e68222. doi: 10.1371/journal.pone.0068222 [published Online First: 2013/07/23]
23. Duan L, Liang R, Wang Z, et al. Lack of association between the G+2044A polymorphism of interleukin-13 gene and chronic obstructive pulmonary disease: a meta-analysis. *Molecular biology reports* 2014;41(9):6297-303. doi: 10.1007/s11033-014-3512-z [published Online First: 2014/07/06]
24. Xie ZK, Huang QP, Huang J, et al. Association between the IL1B, IL1RN polymorphisms and COPD risk: a meta-analysis. *Scientific reports* 2014;4:6202.
25. Hersh C, Dahl M, Ly N, et al. Chronic obstructive pulmonary disease in α1-antitrypsin PI MZ heterozygotes: a meta-analysis. *Thorax* 2004;59(10):843-49.
26. Dahl M, Hersh C, Ly N, et al. The protease inhibitor PI* S allele and COPD: a meta-analysis. *European Respiratory Journal* 2005;26(1):67-76.
27. Ma T, Liu X, Liu Z. Functional polymorphisms in surfactant protein genes and chronic obstructive pulmonary disease risk: a meta-analysis. *Genetic testing and molecular biomarkers* 2013;17(12):910-7. doi: 10.1089/gtmb.2013.0308 [published Online First: 2013/10/08]
28. Zhang L, Gu H, Gu Y, et al. Association between TNF-alpha -308 G/A polymorphism and COPD susceptibility: a meta-analysis update. *International journal of chronic obstructive pulmonary disease* 2016;11:1367-79. doi: 10.2147/copd.s105394 [published Online First: 2016/07/07]
29. Cui K, Ge XY, Ma HL. Association of -238G/A and -863C/A polymorphisms in the TNF-alpha gene with chronic obstructive pulmonary disease based on a meta-analysis. *Genetics and Molecular Research* 2013;12(4):4981-89.
30. Jayes L, Haslam PL, Gratziou CG, et al. SmokeHaz Systematic Reviews and Meta-analyses of the Effects of Smoking on Respiratory Health. *Chest* 2016;150(1):164-79. doi: 10.1016/j.chest.2016.03.060
31. Kamal R, Srivastava AK, Kesavachandran CN. Meta-analysis approach to study the prevalence of chronic obstructive pulmonary disease among current, former and non-smokers. 2015
32. Wang BY, Xiao D, Wang C. Smoking and chronic obstructive pulmonary disease in Chinese population: a meta-analysis. *Clinical Respiratory Journal* 2015;9(2):165-75. doi: 10.1111/crj.12118

33. Wang Q, Li S, Li H, et al. Relative contribution of genetic, individual, and self-efficacy factors to smoking cessation in a Chinese rural population. *The American journal on addictions* 2017;26(2):161-66.
34. Fischer F, Kraemer A. Meta-analysis of the association between second-hand smoke exposure and ischaemic heart diseases, COPD and stroke. *Bmc Public Health* 2015;15 doi: 10.1186/s12889-015-2489-4
35. Zheng PF, Shu L, Si CJ, et al. Dietary Patterns and Chronic Obstructive Pulmonary Disease: A Meta-analysis. *Copd-Journal of Chronic Obstructive Pulmonary Disease* 2016;13(4):515-22. doi: 10.3109/15412555.2015.1098606
36. Po JYT, FitzGerald JM, Carlsten C. Respiratory disease associated with solid biomass fuel exposure in rural women and children: systematic review and meta-analysis. *Thorax* 2011;66(3):232-39. doi: 10.1136/thx.2010.147884
37. Hu G, Zhou Y, Tian J, et al. Risk of COPD from exposure to biomass smoke: a metaanalysis. *Chest Journal* 2010;138(1):20-31.
38. Sana A, Somda SM, Meda N, et al. Chronic obstructive pulmonary disease associated with biomass fuel use in women: a systematic review and meta-analysis. *BMJ Open Respiratory Research* 2018;5(1):e000246.
39. Alif SM, Dharmage SC, Bowatte G, et al. Occupational exposure and risk of chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Expert Review of Respiratory Medicine* 2016;10(8):861-72. doi: 10.1080/17476348.2016.1190274
40. Ryu JY, Sunwoo YE, Lee SY, et al. Chronic Obstructive Pulmonary Disease (COPD) and Vapors, Gases, Dusts, or Fumes (VGDF): A Meta-analysis. *Copd-Journal of Chronic Obstructive Pulmonary Disease* 2015;12(4):374-80. doi: 10.3109/15412555.2014.949000
41. Waziry R, Jawad M, Ballout RA, et al. The effects of waterpipe tobacco smoking on health outcomes: an updated systematic review and meta-analysis. *International journal of epidemiology* 2017;46(1):32-43.
42. Ortega VE. Picking the Right Fruit: Intersecting COPD GWAS Discoveries with Epigenetics. *American journal of respiratory and critical care medicine* 2018(ja)
43. Ewa B, Danuta M-Š. Polycyclic aromatic hydrocarbons and PAH-related DNA adducts. *Journal of applied genetics* 2017;58(3):321-30.
44. Vibhuti A, Arif E, Mishra A, et al. CYP1A1, CYP1A2 and CYBA gene polymorphisms associated with oxidative stress in COPD. *Clinica chimica acta; international journal of clinical chemistry* 2010;411(7-8):474-80. doi: 10.1016/j.cca.2009.12.018 [published Online First: 2010/01/19]
45. Srogi K. Monitoring of environmental exposure to polycyclic aromatic hydrocarbons: a review. *Environmental Chemistry Letters* 2007;5(4):169-95.
46. Hardin M, Cho MH, Sharma SS, et al. Sex-based Genetic Association Study Identifies CELSR1 as a Possible COPD Risk Locus Among Women. *American journal of respiratory cell and molecular biology* 2016 doi: 10.1165/rcmb.2016-0172OC [published Online First: 2016/11/18]
47. Cui K, Ge X, Ma H. Four SNPs in the CHRNA3/5 alpha-neuronal nicotinic acetylcholine receptor subunit locus are associated with COPD risk based on meta-analyses. *PloS one* 2014;9(7):e102324. doi: 10.1371/journal.pone.0102324 [published Online First: 2014/07/23]
48. Cho MH, McDonald ML, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The Lancet Respiratory medicine*

2014;2(3):214-25. doi: 10.1016/s2213-2600(14)70002-5 [published Online First: 2014/03/14]
49. Sherva R, Wilhelmsen K, Pomerleau CS, et al. Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit alpha 5 (CHRNA5) with smoking status and with 'pleasurable buzz'during early experimentation with smoking. *Addiction* 2008;103(9):1544-52.
50. Mizoguchi H, Yamada K, Mouri A, et al. Role of matrix metalloproteinase and tissue inhibitor of MMP in methamphetamine-induced behavioral sensitization and reward: implications for dopamine receptor down-regulation and dopamine release. *Journal of neurochemistry* 2007;102(5):1548-60.
51. Yamada K. Pro-addictive and anti-addictive factors for drug dependence. 2008
52. Navratilova Z, Kolek V, Petrek M. Matrix metalloproteinases and their inhibitors in chronic obstructive pulmonary disease. *Archivum immunologiae et therapiae experimentalis* 2016;64(3):177-93.
53. Yao H, Rahman I. Current concepts on the role of inflammation in COPD and lung cancer. *Current opinion in pharmacology* 2009;9(4):375-83.
54. Shaykhiev R, Crystal RG. Early events in the pathogenesis of chronic obstructive pulmonary disease. Smoking-induced reprogramming of airway epithelial basal progenitor cells. *Annals of the American Thoracic Society* 2014;11(Supplement 5):S252-S58.
55. Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *Human genetics* 2009;125(2):131-51.
56. Stoller JK, Aboussouan LS. α1-antitrypsin deficiency. *The Lancet* 2005;365(9478):2225-36.
57. Society AT. What is Alpha-1 Antitrypsin Deficiency? ATS Patient Education Series, 2014.
58. Gunes A, Ozbey G, Vural EH, et al. Influence of genetic polymorphisms, smoking, gender and age on CYP1A2 activity in a Turkish population. 2009
59. Hüls A, Ickstadt K, Schikowski T, et al. Detection of gene-environment interactions in the presence of linkage disequilibrium and noise by using genetic risk scores with internal weights from elastic net regression. *BMC genetics* 2017;18(1):55.
60. de Boer A, van de Worp WR, Hageman GJ, et al. The effect of dietary components on inflammatory lung diseases–a literature review. *International Journal of Food Sciences and Nutrition* 2017:1-17.
61. Svanes C, Omenaas E, Jarvis D, et al. Parental smoking in childhood and adult obstructive lung disease: results from the European Community Respiratory Health Survey. *Thorax* 2004;59(4):295-302.
62. Drummond D, Baravalle-Einaudi M, Lezmi G, et al. Combined effects of in utero and adolescent tobacco smoke exposure on lung function in C57Bl/6J mice. *Environmental health perspectives* 2017;125(3):392.
63. Perret JL, Walters H, Johns D, et al. Mother's smoking and complex lung function of offspring in middle age: A cohort study from childhood. *Respirology* 2016;21(5):911-19.
64. Hopewell S, McDonald S, Clarke MJ, et al. Grey literature in meta-analyses of randomized trials of health care interventions. *The Cochrane Library* 2007
65. Miravitlles M, Calle M, Soler-Cataluna JJ. Clinical phenotypes of COPD: identification, definition and implications for guidelines. *Archivos de Bronconeumología (English Edition)* 2012;48(3):86-98.

66. Tramèr MR, Reynolds DJM, Moore RA, et al. Impact of covert duplicate publication on meta-analysis: a case study. *Bmj* 1997;315(7109):635-40.
67. Kutmon M, Lotia S, Evelo CT, et al. WikiPathways App for Cytoscape: Making biological pathways amenable to network analysis and visualization. *F1000Research* 2014;3
68. Kao PY, Leung KH, Chan LW, et al. Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions. *Biochimica et Biophysica Acta (BBA)-General Subjects* 2016

Supplementary Materials

Full search strategy

Genetic factors

Pubmed: (pulmonary disease, chronic obstructive[MeSH Terms] OR ("pulmonary"[All Fields] AND "disease"[All Fields] AND "chronic"[All Fields] AND "obstructive"[All Fields])OR "chronic obstructive pulmonary disease"[All Fields] OR "copd"[All Fields] OR pulmonary emphysema[MeSH Terms] OR "emphysema"[All Fields] OR "pulmonary emphysema"[All Fields] OR "emphysema"[MeSH Terms] OR bronchitis, chronic[MeSH Terms] OR "bronchitis"[All Fields] OR "chronic bronchitis"[All Fields] OR lung diseases, obstructive[MeSH Terms] OR ("lung"[All Fields] AND "diseases"[All Fields] AND "obstructive"[All Fields])OR "obstructive lung diseases"[All Fields] OR ("obstructive"[All Fields] AND "pulmonary"[All Fields] AND "disease"[All Fields])OR "obstructive pulmonary disease"[All Fields])AND (genome-wide association study[MeSH Terms] OR (("genome-wide"[All Fields] OR ("genome"AND "wide"))AND "association"[All Fields] AND "study"[All Fields])OR "gwas"[All Fields] OR candidate gene[All Fields] OR candidate genes[All Fields] OR polymorphism[All Fields] OR polymorphisms[All Fields])

Web of Science: ((pulmonary disease, chronic obstructive)OR (pulmonary AND disease AND chronic AND obstructive)OR chronic obstructive pulmonary disease OR copd OR pulmonary emphysema OR emphysema OR pulmonary emphysema OR emphysema OR bronchitis, chronic OR bronchitis OR chronic bronchitis OR lung diseases, obstructive OR (lung AND diseases AND obstructive OR obstructive lung diseases OR (obstructive AND pulmonary AND disease)OR obstructive pulmonary disease))AND ((genome-wide association study)OR ((genome-wide OR (genome AND wide)AND association AND study)OR gwas OR (candidate gene OR candidate genes)OR (polymorphism OR polymorphisms)))

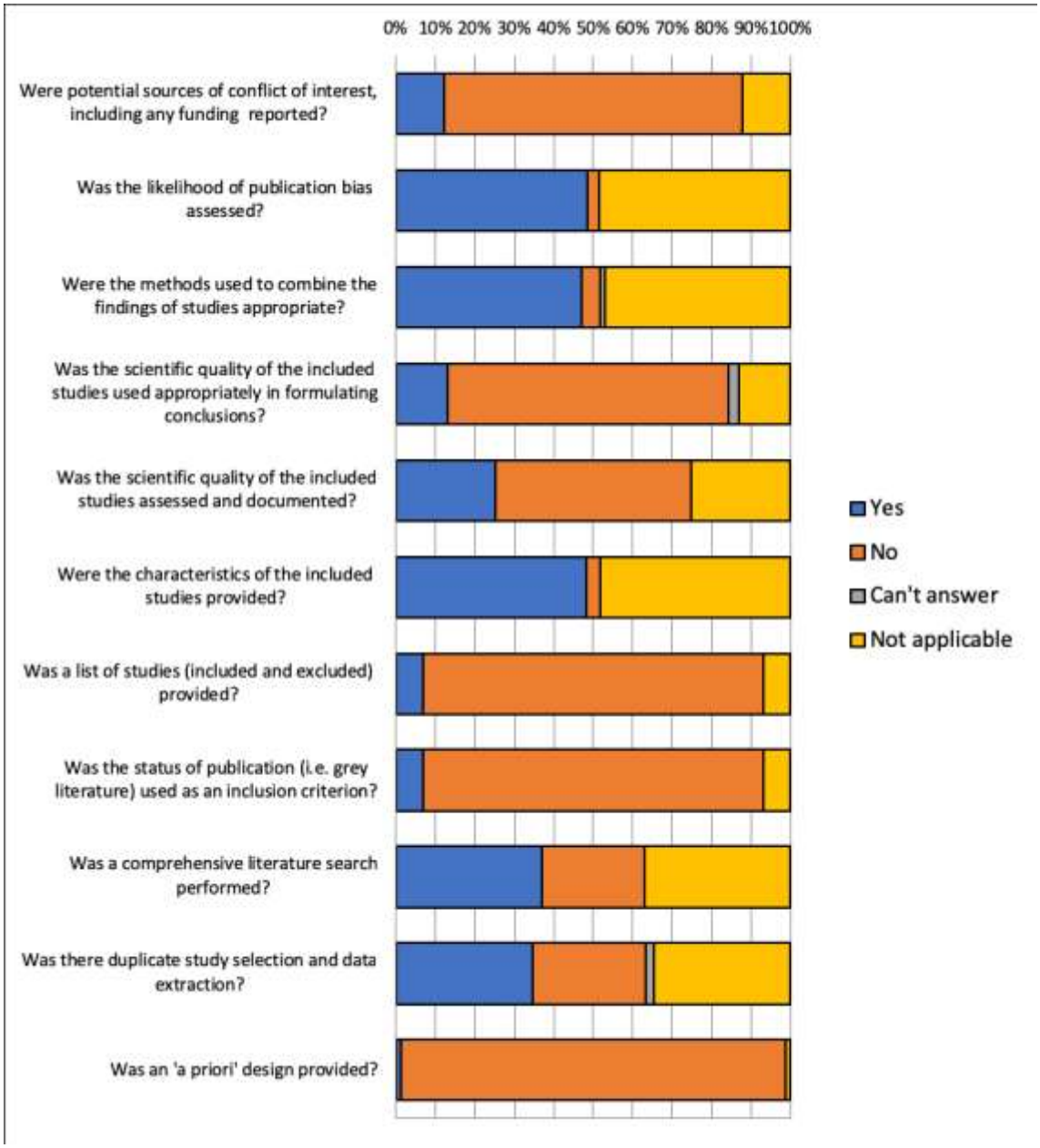
Embase: (((((((pulmonary disease, chronic obstructive OR ("pulmonary" AND "disease" AND "chronic" AND "obstructive")OR "chronic obstructive pulmonary disease" OR "copd" OR pulmonary emphysema OR "emphysema" OR "pulmonary emphysema" OR "emphysema" OR bronchitis, chronic OR "bronchitis" OR "chronic bronchitis" OR lung diseases, obstructive OR ("lung" AND "diseases" AND "obstructive" OR "obstructive lung diseases" OR ("obstructive" AND "pulmonary" AND "disease")OR "obstructive pulmonary disease"))))))AND ((genome-wide association study OR ((("genome-wide" OR ("genome"AND "wide"))AND "association" AND "study")OR "gwas" OR candidate gene OR candidate genes)OR (polymorphism OR polymorphisms))))){No Related Terms}

Lifestyle and environmental factors

PubMed: ("Pulmonary Disease, Chronic Obstructive"[Mesh] OR "Pulmonary Emphysema"[Mesh] OR "Bronchitis, Chronic"[Mesh])AND ("Diet"[Mesh] OR "Exercise"[Mesh] OR "Smoking"[Mesh] OR "Environment"[Mesh] OR "Occupations"[Mesh] OR "Ethanol"[Mesh])AND ("Meta-Analysis"[Publication Type] OR "pooled analysis”)

Web of Science: ("Chronic Obstructive Pulmonary Disease" OR "COPD" OR "emphysema" OR "chronic bronchitis") AND (diet OR dietary OR nutrition* OR exercise OR "physical activity" OR smoking OR tobacco OR environment* OR occupation* OR alcohol OR alcoholic)AND ("Meta-Analysis" OR "pooled analysis")

Embase: (chronic obstructive lung disease/ OR emphysema/ OR chronic bronchitis/)AND (diet/ OR dietary intake/ OR nutrition/ OR exercise/ OR physical activity/ OR smoking/ OR tobacco/ OR tobacco consumption/ OR environment/ OR occupation/ OR alcohol/ OR alcohol consumption/)AND (meta analysis/ OR "pooled analysis")



Supplementary **Figure 5**: Amstar ratings

Supplementary Tables

Supplementary Table 6

Author (year)	Genetic factors not included	Genetic factors included as most comprehensive	Genetic factors included in MMA	COPD diagnosis criteria	Number of included studies	Total cases / controls	Included ethnicities	Evidence for publication bias?	Smoking behavior	AMSTAR quality score
Aierken (2014)[1]	ADAM33 (rs2280091, rs3918396, rs528557)	-	-	GOLD, $FEV_1/FVC<0.7$ and pp-$FEV_1<75\%$	10	2139/3765	C, A	No	UMNR	5/11
An (2016)[2]	EPHX1 (rs1051740, rs2234922)	-	-	Clinical criteria	19	7699/41935	C, E-A	No	UMNR	5/11
Brøgger (2006)[3]	EPHX1 (rs1051740, rs2234922), TNF (rs1800629)	-	-	Post-BD pp-$FEV_1<80\%$ and FEV1/ FVC<0.7	16	500/725 for rs1051740, 829/1282 for rs2234922, 881/1232 for rs1800629	C, A	NR	UMNR	3/11
Castaldi (2010)[4]	ACE (I/D), ADRB2 (rs1042713, rs1042714), EPHX1 (rs2234922, rs1051740), GSTP1 (rs1695), IL1B (rs1143627, rs16944), IL6 (rs1800795), IL13 (rs20541, rs1800925), MMP9 (rs3918242), TGFB1 (rs1800469, rs1800470), TNF (rs1800610, rs1799964, rs361525, rs1800629, rs1799724, rs1800630), GSTM1 (Null/Wt), GSTT1 (Null/Wt), SOD3 (rs1799895)	IL4 (rs2070874), LTA (rs909253), SERPINA3 (rs4934), TIMP2 (rs2277698)	-	pp-$FEV_1<80\%$, pp-$FEV_1<70\%$, FEV1/FVC<0.7, not reported	108	11401/23775	NR, included at least C and A	No	UMNR	3/11
Chen (2015)[5]	VDBP	-	-	CMA (1997, 2002); ATS; FEV1/FVC<0.7 and post-BD pp-	8	809/1407	C, A	Yes	UMNR	6/11

				FEV_1<80%; post-BD FEV1/FVC<0.7						
Chen (2013)[6]	-	IL13 (rs1800925)	-	Not reported	8	1319/831	C, A, Ar	No	UMNR	4/11
Chen (2013)[7]	MMP1 (rs1799750)	MMP9 (rs17576)	-	Not reported	8 for MMP1, 11 for MMP9	2174/2275 for MMP1, 1638/1726 for MMP9	C, A	No	UMNR	4/11
Cho (2010)[8]	CHRNA3 (rs1051730), HHIP (rs13118928), HYKK (rs8034191), IREB2 (rs13180)	HHIP (rs1828591), IREB2 (rs1062980), LOC105377462 (rs720485)	-	pp-FEV_1< 80% and FEV_1/ FVC<0.7; pp-FEV_1<45% and emphysema-CT; GOLD Stage ≥ II; pp-FEV_1< 60% and pp-FEV1/FVC<90%	4	3442 / 1884	C, AA	N/A	UMNR	2/4
Cho (2012)[9]	HHIP (rs13141641)	FAM13A (rs1964516, rs7671167), RAB4B (rs7937, rs2604894), HHIP (rs13118928), IREB2 (rs11858836, rs13180)	-	pp-FEV_1<80% and FEV_1/ FVC<0.7; pp-FEV_1<45% and emphysema-CT; GOLD Stage ≥ II; pp-FEV_1<60% and pp-FEV_1/FVC<90%	4 (no systematic literature search)	3499/1922	C	N/A	UMNR	3/4
Cho (2014)[10]	HHIP (rs13141641)	FAM13A (rs4416442), CHRNA3 (rs12914385), MMP3/MMP12 (rs626750), TGFB2 (rs4846480)	-	Post-BD pp-FEV_1<80% and pp-FEV_1/ FVC<70%	4 (no systematic literature search)	6633/5704	C, AA	N/A	UMNR	2/4
Cui (2014)[11]	-	CHRNA3 (rs1051730, rs6495309), CHRNA5 (rs16969968), HYKK (rs8034191)	CHRNA5 (rs16969968)	Post-BD pp-FEV_1<80% and pp-FEV_1/ FVC<70%; ICD-8 491 to 492 and ICD-10 J41 to J44; pp-FEV_1/ FVC≤70%; post-BD pp-FEV_1/ FVC<70%; pre-BD pp-FEV_1/FVC< 70%, and pp-FEV_1<80%	14	10466/39054 for rs1051730, 2652/2565 for rs8034191, 1977/2131 for rs6495309, 1996/6463 for rs16969968	Unclear, at least A, C, AA	Yes	UMNR	6/11

Cui (2013)[12]	-	TNF (rs361525, rs1800630)	-	ATS, GOLD, CMA (2002)	6	NR	C, A	No	UMNR	5/11
Cui (2015)[13]	-	TNF (rs80267959)	-	GOLD, CMA (2002, 2007), ATS	10	1184/1439	C, A	No	UMNR	6/11
Dahl (2005)[14]	-	SERPINA1 (PI SM, PI SZ)	-	Physician diagnosis, spirometry	6 for PI SZ and 17 for PI MS		C	No	UMNR	6/11
Ding (2018)[15]	-	GSTM1, GSTT1	-	GOLD	37	4674 / 5006	C, A, Af	N/A	UMNR	4/11
Du (2016)[16]	-	-	IREB2 (rs2568494) with Hardin (2016)	Physician diagnosis	4	1513/1480	C, A, In	No	UMNR	6/11
Duan (2013)[17]	-	IL13 (rs20541)	-	pp-FEV_1<70%; pp-FEV_1/FVC<70%; Post-BD pp-FEV_1<80 % and pp-FEV_1/FVC<70%	16	1213/801	C, A, Af	No	UMNR	4/11
Durme, van (2010)[18]	HHIP (rs13118928)	-	-	pp-FEV_1/FVC<70% and pp-FEV_1<80% or physician diagnosis for definite cases; pp-FEV_1/FVC <0.7 and pp-FEV_1≥80% for 'probable cases'	3 (no systematic literature search)	NR	C	N/A	UMNR	0/4
Gong (2011)[19]	TGFB1 (rs1800469, rs1800470)	-	-	GOLD	7 for rs1800469 and 8 for rs1800470	1508/2608	C, A	No	UMNR	5/11
Hardin (2016)[20]	RIN3 (rs754388)	AGPHD1 (rs8042849, rs9788721), CELSR1 (rs56344079, rs7286446, rs9615358, rs9615973, rs9615981, rs9615982), CHRNA3 (rs1051730, rs114205691, rs12914385, rs138544659, rs141518190, rs146009840, rs147144681,	IREB2 (rs2568494) with Du (2016)	GOLD stage ≥ II	3 (no systematic literature search)	6260/5269	C, AA	N/A	Current and former smokers	1/4

		rs147499554, rs4887067, rs55676755, rs56077333, rs8192482), CHRNA5 (rs11633958, rs17486278, rs140330585, rs16969968, rs17486195, rs190065944, rs2036527, rs55853698, rs56390833, rs7172118, rs7180002, rs72740955, rs72740964, rs951266), CHRNB4 (rs17487223, rs55988292, rs72743158), EEFSEC (rs2811416), EPB41L4 A-AS1 (rs66669542), FAM13A (rs10021465, rs6837671, rs1812329, rs2013701, rs2045517, rs28455964, rs2869966, rs2869967, rs2904259, rs3846287, rs3857043, rs4416442, rs4693980, rs6830970, rs76273989, rs7671167, rs7671261, rs7674369, rs7682317, rs7682431), FAR2 (rs7294481), FOXO1 (rs78372177, rs75700692), GEMIN4 (rs11652959), GYPA (rs13105210, rs4835177), IREB2 (rs2656052, rs2656065, rs2938670), KBTBD12 (rs17282209), PSMA4 (rs58365910), RIN3 (rs1075472, rs72699855), VPS53 (rs11247558, rs11656538, rs34001232, rs34469205, rs34729304, rs35716682, rs4968100, rs4968102)								
Hersh (2004)[21]	-	SERPINA1 (PI MZ, PI ZZ, PI MM)	-	Physician diagnosis; $FEV_1/FVC<0.72$	15	275/2579 For PI MZ/PI MM; 1454/13 163	C	No	UMNR	7/11

				and pp-PEFR<80%; PFTs and blood gases; pp-MMEF<50%; pp-TLCO<80%; pp-FEV_1<70%; FEV_1/FVC<0.5; plethysmography; pp-FEV_1<80% and pp-FEV_1/FVC<70%; FEV_1<2 SD below percent predicted		For PI MZ/PI MM				
Hu (2008)[22]	EPHX1 (rs1051740, rs2234922)	-	-	Chest CT; pp-FEV_1/FVC<70%; pp-FEV_1<80% and pp-FEV_1/FVC<70%; ATS; Questionnaire; pp-FEV_1/FVC<70%; CMA (1997, 2002)	16	1847/2455	C, A	No	UMNR	4/11
Hu (2008)[23]	GSTM1 (Null/Wt), GSTT1 (Null/Wt)	-	-	Chest CT, GOLD, ATS, ETS; Physician diagnosis; Symptom and signs and pp-FEV_1<80% and pp-FEV_1/FVC<70%; pp-FEV_1⩽45%, predicted hyperinflation on PFT, bilateral emphysema on chest HRCT; Chronic bronchitis according to ICD-10; Chronic bronchitis *; COPD and	12 for GSTM1, 8 for GSTT1	1697/1867 for GSTM1 and 1320/1435 for GSTT1	C, A, Tr	No	UMNR	5/11

				Emphysema **; Pathology Emphysema; Chronic Bronchitis						
Hobbs (2016)[24]	AGER (rs2070600)	AKD1 (rs10499052), CRAMP1L (rs61746451), FAM208B (rs41290259), IL27 (rs181206), MICAL1 (rs59056467), MMP3 (rs679620), SERPINA1 (rs28929474), TIRAP (rs8177374)	CHRNA5 (rs16969968)	FEV_1/FVC<0.7 and pp-FEV_1<80%; severe COPD: FEV_1/FVC<0.7 and pp-FEV_1<50%	5	6004/6191	Unclear, at least A, C, AA	N/A	UMNR	1/4
Hobbs (2017)[25]	-	ADAM19 (rs113897301), ADGRG6 (rs9399401), AGER (rs2070600), ARMC2 (rs2806356), CCDC101 (rs17707300), CFDP1 (rs7186831), CHRNA5 (rs17486278), CYP2A6 (rs12459249), DSP (rs2076295), EEFSEC (rs2955083), FAM13A (rs6837671), GSTCD (rs11727735), HHIP (rs13141641), HTR4 (rs7733088), MTCL1 (rs647097), PID1 (rs16825267), RARB (rs1529672), RIN3 (rs754388), SFTPD (rs721917), TET2 (rs2047409), TGFB2 (rs10429950), THSD4 (rs1441358)	-	GOLD; GOLD stage ≥ II; pp-FEV_1<40%; pp-FEV_1<60% and FEV_1/FVC<0.9; pp-FEV_1< 50%; not reported	26	15256/47936	C, AA, A, H	NR	UMNR	2/4
Ji (2017)[26]		IL6 (rs1800796)		Not reported	3	413/596	C, A	No	UMNR	4/11
Jiang (2016)[27]	MMP1 (rs1799750)	MMP3 (rs3025058), MMP9 (rs3918242)	-	Not reported	12	650/627 for MMP1, 522/520 for MMP3, 1067/1091 for MMP9	C, A	Yes	UMNR	6/11

Kang (2016)[28]	-	ACE (I/D)	-	Not reported	14	977/1092	C, A	No	UMNR	4/11
Lee (2011)[29]	EPHX1 (rs1051740)	EPHX1 (rs2234922)	-	pp-FEV_1/FVC<70%; GOLD	19	7489/42970	C, A	NR	UMNR	3/4
Li (2014)[30]	ADAM33 (rs528557, rs2280091)	-	-	Not reported	6	838/831 for rs528557, 915/912 for rs2280091	Ch	No	Current and former smokers	4/11
Li (2013)[31]	EPHX1 (rs2234922)	EPHX1 (rs1051740)	-	Not reported	25	5186/24510	C, A	Yes	UMNR	4/11
Li (2013)[32]	ACE (I/D)	-	-	ATS; spirometry; GOLD	8	574/787	C, A	No	UMNR	5/11
Li (2012)[33]	ACE (I/D)	-	-	Not reported	10	710/862	C, A	No	UMNR	3/11
Liao (2017)[34]	-	TGF-β1 (rs1800470, rs1800469, rs2241712, rs6957, rs2241718)	-	Not reported	9 for rs1800470, 9 for rs1800469, 4 for rs2241712, 5 for rs6957, 3 for rs2241718	2451/3247 for rs1800470, 2405/3338 for rs1800469, 807/906 for rs2241712, 1732/2751 for rs6957, 609/746 for rs2241718	C, A	No	UMNR	5/11
Ma (2013)[35]	-	SP-A/B/D	-	pp-FEV_1/FVC<70% or pp-FEV_1/FVC<80% with medication use	7	1274/1482	C, A	No	UMNR	6/11
Nedeljkovic (2018)[36]	-	SLC22A11 (rs141159367), MTL5 (rs146043252)	-	pp-FEV_1/FVC<70%	9	9888 / 27428	C, A, Af,	N/A	UMNR	2/4
Nielsen (2017)[37]	-	ADRB2 (rs1800888, rs1042713, rs1042714)	-	Post-BD pp-FEV_1/FVC<70% or GOLD	3 for rs1800888 heterozygous, 2 for rs1800888 homozygous, 12 for rs1042713	303/1577 for rs1800888 heterozygous, 9/11 for rs1800888 homozygous, 1059/3855 for rs1042713	C, A, Af	No	UMNR	3/11

					homozygous, 12 for rs1042713 heterozygous, 10 for rs1042714 homozygous, 9 for rs1042714 heterozygous	homozygous, 831/3182 for rs1042713 heterozygous, 959/3939 for rs1042714 homozygous, 394/1627 for rs1042714 heterozygous				
Niu (2012)[38]	ADRB2 (rs1042713, rs1042714)	-	-	ATS, ERS; GOLD	20	912/1145 for rs1042713, 994/1183 for rs1042714	C, A	No	UMNR	3/11
Smolonska (2009)[39 40]	EPHX1 (rs1051740, rs2234922), GSTM1 (Null/Wt), GSTT1 (Null/Wt), GSTP1 (rs1695), IL1B (rs16944, rs1143627, rs1143634), IL1RN (rs2234663/VNTR), IL6 (rs1800795), MMP9 (rs3918242), TGFB1 (rs1800470, rs1800469, rs2241712, rs6957)	SOD2 (rs4880), SOD3 (rs1799895), TNF (rs180061)	-	$FEV_1/FVC<0.7$; pp-$FEV_1<80\%$ and $FEV_1/FVC<0.7$; pp-$FEV_1<60\%$ and $FEV_1/FVC<90\%$; pp-$FEV_1<45\%$; not reported	69	3768/6078 for rs1051740, 4082/6541 for rs2234922, 2787/6393 for GSTM1 Null/Wt, 2189/4675 for rs1695, 1280/1422 for GSTT1 Null/Wt, 1798/1979 for rs16944, 2787/6393 for rs1143627, 740/712 for rs1143634, 318/385 for rs2234663/VNTR, 1084/1107 for rs1800795, 828/1003 for rs3918242, 807/2389 for rs4880, 1392/9237 for rs1799895, 747/932 for	C, A	NR	UMNR	4/11

						rs2241712, 4580/4670 for rs1800470, 1246/3010 for rs1800469, 963/2492 for rs6957, 2010/1864 for rs361525, 2005/1891 for rs180061, 4580/4670 for rs1800629				
Tian (2015)[41]	-	SERPINE2 (rs3795879)	-	ATS, ERS	5	3034/3068	C, A	No	UMNR	5/11
Wain (2017)[42]	-	ABLIM3 (rs3839234), ADAM19 (rs1990950), ADGRG6 (rs7753012, rs148274477), AGER (rs2070600), AHNAK (rs2509961), ARL15 (rs2441026), ARMC2 (rs2768551), ASTN2 (rs803923), BMP6 (rs6924424), C1GALT1 (rs10246303), C5orf56 (rs7713065), CACNA2D3/WNT5A (rs1458979), CASC20/BMP2 (rs6140050), CCDC91 (rs2348418), CDC123 (rs7090277), CDC7/TGFBR3 (rs1192404), CFDP1 (rs3743609), CHRM3 (rs6688537), CISD3 (rs11658500), CYFIP2 (rs10515750), DNLZ (rs10870202), EEFSEC (rs2811415), EFCAB5 (rs59835752), EFEMP1 (rs1430193), ENSA (rs6681426), FAM13A	-	Not reported	14	20086/215630	C	N/A	UMNR	2/4

	(rs2045517, rs13110699), FGD6 (rs113745635), GLIS3 (rs7872188), GSTCD (rs10516526), HDAC4 (rs12477314), HLA-DQB1 (rs114544105), HLA-DQB1/HLA-DQA2 (rs34864796, rs114229351), HSD17B12 (rs4237643), HTR4 (rs7715901), ITGA1 (rs1551943), KANSL1 (rs35524223), KCNJ2 (rs6501431), KCNQ5 (rs141651520), KCNS3/RDH14 (rs62126408), LINC00310/KCNE2 (rs2834440), LINC01467/LINC00911 (rs1698268), LOC102723639 (rs35506), LOC105369591 (rs567508), LOC105372926 (rs4328080), LOC105377462 (rs138641402), LOC107984427 (rs145729347), LOC107984437 (rs10850377), LOC389602/LOC285889 (rs12698403), LRMDA (rs2637254), LRP1 (rs11172113), LST1 (rs28986170), LTBP4 (rs113473882), MECOM (rs1344555), MECOM/LOC100507661 (rs56341938), MFAP2 (rs2284746), MGA (rs72724130), MICAL3 (rs11704827), MMP15 (rs12447804), MN1

		(rs134041), MN1 (rs2283847), MSRB3 (rs1494502), MYPN (rs7095607), NCR3/AIF1 (rs2857595), NPNT (rs34712979), PABPC4 (rs17513135), PRDM11 (rs2863171), PTCH1 (rs16909859), QSOX2 (rs10858246), RARB (rs1529672), RIN3 (rs117068593), RSRC1 (rs1595029), SH3GL3 (rs66650179), SNRPF (rs12820313), SPAG17/TBX15 (rs200154334), SPATA9 (rs153916), SPHKAP/PID1 (rs10498230), SUCLG2 (rs1490265), SVIL/JCAD (rs3847402), TARS/LOC340113 (rs91731), TEKT5 (rs12149828), TGFB2/MIR548F3 (rs993925), TGFBR3 (rs12140637), THSD4 (rs12591467), THSD4 (rs10851839), TNS1 (rs2571445), TRAF3IP1/ASB1 (rs61332075), TRIP11 (rs7155279), TSEN54/CASKIN2 (rs7218675), WWOX (rs1079572), ZGPAT (rs72448466), ZKSCAN1 (rs72615157)								
Wang (2015)[43]	-	CYP1A1 (rs4646903, rs1048943)	-	Not reported	7	1050/1202	C, A	No	UMNR	5/11
Wang (2015)[44]	-	IL1B (rs16944)	-	Not reported	13	1692/2009	C, A	No	UMNR	4/11

Wang (2015)[45]	-	VDBP	-	Not reported	12	1485/1659	C, A	No	UMNR	4/11
Wang (2015)[46]	-	ADRB2 (rs1042713)	-	Not reported	11	1128/1182	C, A	No	UMNR	7/11
Xiao (2014)[47]	IL1B (rs1143627)	-	-	Not reported	6	764/879	C, A	No	UMNR	4/11
Xiao (2015)[48]	VDBP	-	-	Not reported	7	734/741	C, A	No	UMNR	4/11
Xie (2015)[49]	VDBP	-	-	Not reported	12	1190/1747	C, A	No	UMNR	5/11
Xie (2015)[50]	-	IL6 (rs1800795)	-	Not reported	7	1144/1656	C, A	No	UMNR	4/11
Xie (2014)[51]	IL1B (rs16944)	IL1B (rs1143627, rs1143634), IL1RN (rs2234663/VNTR)	-	Chest CT; Symptom and signs and pp-FEV_1<80% and FEV_1/FVC<70%; GOLD; ATS; CMA (2002); ETS; History, clinical examination, radiological exams, ABG and PFT abnormal; Clinical diagnosis; FEV_1 < 45% predicted; hyperinflation on PFT	11	1530/1524	E-A, S-A, Ar	No	UMNR	7/11
Xue (2012)[52]	-	GSTM1 (Null/Wt), GSTT1 (Null/Wt)	-	SSPTS; CMA (1997, 2002); GOLD, ERS; ATS; Symptoms and pp-FEV_1<80% and pp-FEV_1/FVC<70%; Clinical diagnosis	14	1665/1614	C, A	No	Smokers	5/11
Yan (2010)[53]	GSTP1 (rs1695)	-	-	Not reported	10	1140/1263	C, A	No	UMNR	7/11
Yang (2015)[54]	-	GSTP1 (rs1695)	-	Post-BD pp-FEV_1<80% and FEV_1/FVC<0.7;	17	1892/2012	C, A	No	UMNR	5/11

				ICD-8 491-492 ICD-10 J41-J44; pre-BD $FEV_1/FVC<0.7$; post-BD $FEV_1/FVC<0.7$						
Zhang (2011)[55]	CHRNA3 (rs1051730)for COPD	CHRNA3 (rs1051730)for EM	-	Not reported	7	3460/11437 (for COPD and EM)	Unclear, likely C	No	UMNR	4/11
Zhang (2011)[56]	TGFB1 (rs1800469, rs1800470)	-	-	Not reported	6 for rs1800469, 10 for rs1800470	1890/4272 for rs1800469, 1507/2542 for rs1800470	C, A	No	UMNR	2/11
Zhang (2016)[57]	-	TNF (rs1800629)	-	GOLD; $FEV_1/FVC<70\%$ and pp-$FEV_1<75\%$	38	3951/5110	Non-A, A	No	UMNR	6/11
Zhang (2014)[58]	ADAM33 (rs511898, rs3918396, rs2280091, rs528557, rs2787094, rs612709)	-	-	CMA (2002); ATS; GOLD, LAA score on chest CT-scans, LAA>8.0; pp-$FEV_1<45\%$ and hyperinflation and EM diagnosis with HRCT; Not reported	12	2630/4376	C, A	No	UMNR	3/11
Zhou (2013)[59]	MMP9 (rs3918242)	MMP1 (rs1799750), MMP3 (rs35068180), MMP12 (rs2276109)	-	GOLD stage ≥ II	21	4184/5716	C, A	No	UMNR	5/11
Zhou (2015)[60]	-	ADAM33 (rs2787094, rs2280089, rs2280090, rs2280091, rs528557, rs3918396, rs612709, rs511898, rs597980)	-	CMA; GOLD; clinical diagnosis; chest CT and spirometry	13	2644/4804	C, A	Yes	UMNR	4/11
Zhou (2017)[61]		HMOX1 (length polymorphism)			7	712/891	C, A	Yes, for M vs. S+L	UMNR	6/11

Studies included for genetic factors

Legend. *Abbreviations*: N/A = Not applicable, NR = Not Reported, UMNR = Unclear/mixed/not reported; Outcomes: COPD = Chronic Obstructive Pulmonary Disease; Ethnicities: AA = African American, A = Asian, Af = African, Ar = Arabian, C = Caucasian, Ch = Chinese, E-A = East Asian, H = Hispanic, I = Indian, S-A = South Asian, Tr = Turkish; Criteria for COPD diagnosis: ABG = arterial blood gas, ATS = Guidelines of The American Thoracic Society criteria; CMA = Chinese Medicine Association criteria, CT = computed tomography, ERS = European Respiratory Society criteria, FEV_1 = percent-predicted forced expiratory volume in one second, FEV_1/FVC = Forced expiratory volume in one second (FEV_1)to forced vital capacity (FVC)ratio, GOLD = Global Initiative for Chronic Obstructive Lung Disease criteria, HRCT = high resolution computed tomography, LAA = Low Attenuation Area; MMEF = maximum mid-expiratory flow, pp-PEFR= percent predicted peak expiratory flow rate, PFT = pulmonary function test, post-BD = post bronchodilator, pp- = percentage predicted, SSPTS = Spanish Society of Pneumology and Thoracic Surgery criteria, TLCO = carbon monoxide transfer factor

Supplementary Table 2

Author (year)	Risk factor not included	Risk factor included as most comprehensive	Risk factor included in MMA	COPD definition	Included study designs (amount if known)	Total cases / controls	Included countries of study	Evidence for publication bias?	Smoking behavior	AMSTAR quality score
Alif (2016)[62]	Occupational exposure to Vapors, Gases, Dusts, or Fumes (VGDF)	-	-	pre-BD FEV_1/FVC<70% and/or FEV_1/FVC<LLN; post-BD FEV_1/FVC<70%; FEV_1/FVC<LLN	COH (1), CS (4)	9 986 total participants	New Zealand, Netherlands, Australia, Spain, Switzerland	No	Adjusted for smoking	3/11
Fischer (2015)[63]	-	Passive smoking	-	Not reported	CC (4), COH (1)	23 094 total participants	Hong Kong, China, USA, Taiwan	NR	Never smokers or adjusted for smoking	5/11
Forey (2011)[64]	-	Active smoking	Active smoking	ATS; GOLD; BTS; ICD, spirometry; clinical diagnosis, clinical symptoms; clinical diagnoses of CB and or EM and spirometry;	COH (39), CC (20), CS (134)	NR	Worldwide	Yes	N/A	4/11
Hu (2010)[65]	Solid fuel smoke (COPD)	-	-	CB and/or FEV_1/FVC≤70%, CB; FEV1/FVC<70%; GOLD; CB; COPD was defined as reversibility < 12%, pp-FEV_1/FVC<88%; CB and/or pp-FEV_1<75%	CC (4), CS (11)	3 719/38 688	Mexico, Colombia, Bolivia, Brazil, India, Nepal, China, Saudi Arabia, Turkey, Spain	Yes	Mixed, stratified by smoking status	8/11
Jayes (2016)[66]	-	-	Active smoking (COPD)	GOLD; Not reported	Prospective COH (21), Retrospective COH (1), NCC (2)	NR	Multiple European, Australia, USA, Japan, Thailand	No	N/A	7/11

Kamal (2015)[67]	-	-	Active smoking (COPD)	Spirometry; Bronchodilator; Physician diagnosis; Questionnaire; Physician diagnosis and Questionnaire	COH, CS	544 536 total participants	Multiple European, Multiple South America, USA, Multiple Asia, Lebanon	Yes	N/A	4/11
Kurmi (2010)[68]	-	Any solid fuel smoke and wood smoke	-	Hospital diagnosis; GOLD; ATS; BMRC; post FEV_1/FVC ratio<70%; post FEV_1/FVC<70% and pp-FEV_1<80 and <12% change in FEV1 pre-BD vs. post-BD; FEV_1/FVC<0.7 and pp-FEV_1<70% and <15% or <250 ml absolute reduction of FEV_1 pre-BD vs. post-BD	Retrospective COH (1), CC (9), CS (12), DS (1)	55 437 total participants COPD	Saudi-Arabia, Columbia, Spain, Mexico, China, Turkey, Nepal, Bolivia, India, Pakistan, Brazil, Iran	No	Adjusted for smoking	4/11
Po (2011)[69]	Solid fuel smoke	-	-	Not reported	CC (3), CS (9)	16 291 total participants	Pakistan, South Africa, Iran, Turkey, India, Colombia, Turkey, China, Nepal	Yes	Mixed, stratified by smoking status	5/11
Ryu (2014)[70]	Occupational exposure to Vapors, Gases, Dusts, or Fumes (VGDF)	-	-	GOLD or FEV_1/FVC<0.7; FEV_1/FVC<LLN; Chronic obstructive bronchitis or symptomatic EM; physician diagnosis	CC (4), CS (7)	26 959 total participants	Norway, USA, UK, South-Africa, China, Spain, Various, Australia	No	Adjusted for smoking	5/11
Sadhra (2017)[71]	-	Occupational exposure to Vapors, Gases, Dusts, or Fumes (VGDF)	-	Physician diagnosis and/or spirometry	COH, CC, CS	Insufficient data	Insufficient data	Yes	Adjusted for smoking	5/11
Sana (2018)[72]	-	Biomass smoke exposure in females	Biomass smoke exposure in both genders with Yang (2017)[73]	ATS, ERS, GOLD; post-BD FEV_1/FVC<LLN	CC (5), CS (9)	1594/19099	Africa, Asia, South America, Middle East, Europe	No	Adjusted for smoking	1/4

Wang (2015)[74]	-	-	Active smoking (COPD)	ICD; in accordance with the guideline for the diagnosis and treatment COPD	COH (3), CC (4), CS (17)	10 708/93 045	China	Yes	N/A	4/11
Waziry (2017)[75]	-	Waterpipe tobacco exposure	-	Symptomatic COPD; ATS criteria or MRC criteria; physician diagnosis; GOLD; Questionnaire	CC (1), CS (6)	9 079	Insufficient data	NR	Adjusted for other forms of tobacco smoking when possible	3/11
Yang (2017)[73]	Active smoking, Passive smoking	Living circumstances, Frequent cooking, Physical labor work	Biomass smoke exposure (COPD)with Sana (2018)[72], Occupational dust exposure (COPD)with Sadhra (2017)[71]	Not reported	CC (19)	6 383/7 510	China	Yes (smoking estimates only)	Not reported or N/A	5/11
Zheng (2016)[76]	-	Dietary pattern	-	Physician-diagnosis; Spirometry; chest radiography or chest CT	COH (7), CC (3), CS (3)	499 085 total participants	USA, Great Britain, Netherlands, Ireland, China, Japan	No	Most but not all adjusted for smoking	5/11

Studies included for lifestyle and environmental risk factors

Abbreviations: N/A = Not applicable, NR = Not Reported; Criteria for COPD diagnosis: ATS = Guidelines of The American Thoracic Society criteria, BMRC = British Medical Research Council, BTS = British Thoracic Society, CB = chronic bronchitis, CT = computed tomography, EM = emphysema, ERS = European Respiratory Society criteria, FEV_1 = percent-predicted forced expiratory volume in one second, FEV_1/FVC = Forced expiratory volume in one second (FEV_1)to forced vital capacity (FVC)ratio, GOLD = Global Initiative for Chronic Obstructive Lung Disease criteria, ICD = International Classification of Diseases, LLN = lower limit of normal, MRC = medical research council, post-BD = post bronchodilator, pre-BD = pre bronchodilator, pp- = percentage predicted, Study designs: COH = Cohort, CC = Case-control, CS = Cross-sectional; DS = Descriptive study.

Supplementary Table 3

Gene	Variant / rs number	Ethnicity	Smoking*	Gender*	Allelic	Recessive	Dominant	Heterozygous	Homozygous	Overdominant	Other	Risk allele
ACE	del/ins(15)	A, C	UMNR	UMNR	1.17 (0.92 - 1.49)	1.36 (0.89 - 2.09)	1.04 (0.84 - 1.30)	0.90 (0.71 - 1.15)	1.30 (0.84 - 2.01)			D (vs. I)
ACE	del/ins(15)	A	UMNR	UMNR	**1.89 (1.47 - 2.43)**	**3.33 (2.19 - 5.05)**	**1.50 (1.02 - 2.21)**	0.96 (0.62 - 1.47)	**3.13 (1.92 - 5.09)**			D (vs. I)
ACE	del/ins(15)	C	UMNR	UMNR	0.93 (0.79 - 1.09)	0.96 (0.71 - 1.29)	0.87 (0.67 - 1.14)	0.88 (0.66 - 1.17)	0.85 (0.62 - 1.17)			D (vs. I)
ADAM33	rs2280090 (T2)(16)	E, A	UMNR	UMNR	0.66 (0.38 - 1.13)	0.50 (0.14 - 1.80)	0.63 (0.37 - 1.09)		0.43 (0.10 - 1.82)			A (vs. G)
ADAM33	rs2280090 (T2)(16)	E	UMNR	UMNR	1.14 (0.65 - 1.99)	**2.29 (1.11 - 4.71)**	1.09 (0.61 - 1.96)		**2.34 (1.13 - 4.82)**			A (vs. G)
ADAM33	rs2280090 (T2)(16)	A	UMNR	UMNR	**0.46 (0.33 - 0.66)**	**0.22 (0.09 - 0.55)**	**0.44 (0.31 - 0.63)**		**0.17 (0.06 - 0.48)**			A (vs. G)
ADAM33	rs2280091 (T1)(16)	E, A	UMNR	UMNR	1.37 (0.94 - 1.97)	**1.76 (1.27 - 2.43)**	1.40 (0.91 - 2.16)		1.82 (0.88 - 3.77)			G (vs. A)
ADAM33	rs2280091 (T1)(16)	E	UMNR	UMNR	0.93 (0.71 - 1.20)	1.11 (0.65 - 1.88)	0.89 (0.66 - 1.20)		1.06 (0.62 - 1.80)			G (vs. A)
ADAM33	rs2280091 (T1)(16)	A	UMNR	UMNR	**2.03 (1.40 - 2.94)**	**2.31 (1.51 - 3.54)**	**2.24 (1.39 - 3.62)**		**3.19 (1.15 - 8.86)**			G (vs. A)
ADAM33	rs612709 (Q-1)(16)	E, A	UMNR	UMNR	**0.60 (0.52 - 0.68)**	**0.55 (0.44 - 0.69)**	**0.54 (0.45 - 0.65)**		**0.34 (0.23 - 0.49)**			A (vs. G)

ADAM33	rs612709 (Q-1)(16)	E	UMNR	UMNR	**0.64 (0.53 - 0.77)**	**0.65 (0.50 - 0.85)**	**0.53 (0.38 - 0.74)**		**0.51 (0.27 - 0.96)**		A (vs. G)
ADAM33	rs612709 (Q-1)(16)	A	UMNR	UMNR	**0.61 (0.42 - 0.89)**	**0.35 (0.22 - 0.57)**	**0.58 (0.36 - 0.93)**		**0.28 (0.17 - 0.46)**		A (vs. G)
CYP1A1	rs1048943 (17)	C, A	UMNR	UMNR	1.19 (0.81 - 1.74)	**2.75 (1.29 - 5.84)**	1.34 (0.84 - 2.13)	**1.39 (1.01 - 1.90)**	**3.23 (1.50 - 6.93)**		C (vs. T)
CYP1A1	rs1048943 (17)	A	UMNR	UMNR			1.32 (0.56 - 3.11)				C (vs. T)
CYP1A1	rs1048943 (17)	C	UMNR	UMNR	1.50 (0.88 - 2.55)	1.67 (0.09 - 30.45)	1.34 (0.63 - 2.88)	1.19 (0.81 - 1.74)	1.72 (0.09 - 34.41)		C (vs. T)
EPHX1	Combination of rs1051740 and rs2234922 (18)	C, A	UMNR	UMNR						**1.77 (1.23 - 2.55)**	Extremely slow (vs. normal enzyme activity)
EPHX1	Combination of rs1051740 and rs2234922 (18)	A	UMNR	UMNR						1.14 (0.84 - 1.54)	Extremely slow (vs. normal enzyme activity)
EPHX1	Combination of rs1051740 and rs2234922 (18)	C	UMNR	UMNR						**2.64 (1.30 - 5.38)**	Extremely slow (vs. normal enzyme activity)
EPHX1	Combination of rs1051740 and rs2234922 (18)	C, A	UMNR	UMNR						**1.44 (1.13 - 1.85)**	Slow (vs. normal enzyme activity)

EPHX1	Combination of rs1051740 and rs2234922 (18)	A	UMNR	UMNR	1.41 (0.90 - 2.19)	Slow (vs. normal enzyme activity)
EPHX1	Combination of rs1051740 and rs2234922 (18)	C	UMNR	UMNR	**1.31 (1.01 - 1.71)**	Slow (vs. normal enzyme activity)
GSTM1	Null/Wt (19)	C, A, Af	UMNR	UMNR	**1.52 (1.31 – 1.77)**	Null (vs. Wt)
GSTM1	Null/Wt (19)	A	UMNR	UMNR	**1.59 (1.29 – 1.96)**	Null (vs. Wt)
GSTM1	Null/Wt (19)	C	UMNR	UMNR	**1.26 (1.02 – 1.55)**	Null (vs. Wt)
GSTM1	Null/Wt (19)	Af	UMNR	UMNR	**2.42 (1.36 – 4.31)**	Null (vs. Wt)
GSTM1	Null/Wt (20)	C, A	S	UMNR	**1.51 (1.17 - 1.95)**	Null (vs. Wt)
GSTM1	Null/Wt (20)	A	S	UMNR	1.39 (0.98 - 1.96)	Null (vs. Wt)
GSTM1	Null/Wt (20)	C	S	UMNR	**1.59 (1.09 - 2.32)**	Null (vs. Wt)
GSTM1	Null/Wt (20)	C, A	S	M	**2.04 (1.11 - 3.76)**	Null (vs. Wt)

GSTM1	Null/Wt (20)	C, A	S	F	**2.74 (1.50 - 5.02)**	Null (vs. Wt)
HMOX1	Length polymorphism (21)	C, A	UMNR	UMNR	0.72 (0.49 - 1.04)	S (vs. M+L)
HMOX1	Length polymorphism (21)	A	UMNR	UMNR	**0.62 (0.40 - 0.96)**	S (vs. M+L)
HMOX1	Length polymorphism (21)	C, A	UMNR	UMNR	0.91 (0.76 - 1.10)	M (vs. S+L)
HMOX1	Length polymorphism (21)	A	UMNR	UMNR	0.95 (0.73 - 1.23)	M (vs. S+L)
HMOX1	Length polymorphism (21)	C, A	UMNR	UMNR	**2.02 (1.31 - 3.11)**	L (vs. S+M)
HMOX1	Length polymorphism (21)	A	UMNR	UMNR	**2.23 (1.68 - 2.95)**	L (vs. S+M)
HMOX1	Length polymorphism (21)	C, A	UMNR	UMNR	**1.82 (1.28 - 2.61)**	Type I (at least one L allele)(vs. Type II (non-L-allele-carrier))
HMOX1	Length polymorphism (21)	A	UMNR	UMNR	**2.02 (1.51 - 2.7)**	Type I (at least one L allele)(vs. Type II (non-L-allele-carrier))

IL6	rs1800796 (22)	C, A	UMNR	UMNR				**0.45 (0.28 - 0.75)**	**0.20 (0.09 - 1.38)**		C (vs. G)
IL13	rs1800925 (23)	C, A, Ar, Uygur	UMNR	UMNR				**1.82 (1.14 - 2.92)†**	**2.02 (1.10 - 3.72)**		T (vs. C/G)
IL13	rs1800925 (23)	A	UMNR	UMNR				2.02 (0.71 - 5.73)†	1.81 (0.23 - 14.46)		T (vs. C/G)
IL13	rs1800925 (23)	C	UMNR	UMNR				1.20 (0.57 - 2.51)†	1.70 (0.68 - 4.24)		T (vs. C/G)
IL13	rs1800925 (23)	Ar	UMNR	UMNR				**2.94 (1.03 - 8.42)†**	**3.05 (1.08 - 8.60)**		T (vs. C/G)
IL13	rs20541 (24)	C, Af, A	UMNR	UMNR	1.12 (0.96 - 1.32)	1.18 (0.97 - 1.44)	0.99 (0.49 - 2.00)			0.85 (0.70 - 1.04)	G (vs. A)
IL13	rs20541 (24)	C	UMNR	UMNR	0.87 (0.67 - 1.14)	0.95 (0.70 - 1.30)	**0.38 (0.14 - 0.98)**			0.91 (0.66 - 1.24)	G (vs. A)
IL13	rs20541 (24)	A	UMNR	UMNR	**1.30 (1.05 - 1.61)**	**1.34 (1.02 - 1.76)**	1.60 (0.99 - 2.58)			0.87 (0.66 - 1.15)	G (vs. A)
IL1RN	rs2234663 (25)	E-A, S-A, Ar	UMNR	UMNR		**2.59 (1.02 - 6.58)**	1.64 (0.99 - 2.73)		**3.16 (1.23 - 8.13)**		2 (vs. L)
IL1RN	rs2234663 (25)	E-A	UMNR	UMNR		2.60 (0.93 - 7.31)	1.48 (0.77 - 2.83)		**3.20 (1.13 - 9.12)**		2 (vs. L)
SERPINA1	PI SZ (27)	C	UMNR	UMNR	**3.26 (1.24 - 8.57)**						SZ (vs. MM)

SERPINA1	PI MZ (26)	C	UMNR	UMNR	**2.31 (1.60-3.35)**					Z (vs. M)
SERPINA1	PI MZ (26)	C	S	UMNR	1.61 (0.92 - 2.81)					Z (vs. M)
SFTPA1/SFTPA2/SFTPB/SFTPD combined	Combination of rs1059046, rs1136451, rs4253527, rs1130866, rs2077079, rs1051246, rs2245121, rs2255601, rs3088308, rs6413520, rs721917 and rs911887 (28)	C, A	UMNR	UMNR	**1.21 (1.01 - 1.45)**		1.30 (0.94 - 1.80)			W (vs. M)[$]
SFTPA1/SFTPA2/SFTPB/SFTPD combined	Combination of rs1059046, rs1136451, rs4253527, rs1130866, rs2077079, rs1051246, rs2245121, rs2255601, rs3088308, rs6413520, rs721917 and rs911887 (28)	A	UMNR	UMNR	**1.43 (1.15 - 1.78)**	**1.48 (1.11 - 1.97)**	**1.97 (1.38 - 2.81)**	**1.39 (1.04 - 1.85)**	**2.26 (1.56 - 3.28)**	W (vs. M)[$]
SFTPA1/SFTPA2/SFTPB/SFTPD combined	Combination of rs1059046, rs1136451, rs4253527, rs1130866, rs2077079, rs1051246, rs2245121, rs2255601,	C	UMNR	UMNR	0.99 (0.79 - 1.24)	1.00 (0.71 - 1.41)	0.85 (0.59 - 1.22)	1.02 (0.76 - 1.37)	0.82 (0.50 - 1.36)	W (vs. M)[$]

	rs3088308, rs6413520, rs721917 and rs911887 (28)									
SFTPA1/SFTPA2	Combination of rs1059046, rs1136451 and rs4253527 (28)	C, A	UMNR	UMNR	**1.53 (1.14 - 2.05)**	**1.66 (1.17 - 2.35)**	**1.65 (1.02 - 2.69)**	**1.59 (1.13 - 2.22)**	**2.06 (1.24 - 3.42)**	W (vs. M)[$]
TNF	rs1800629 (29)	A, Non-A	UMNR	UMNR	**1.56 (1.29 - 1.89)**			**1.51 (1.26 - 1.81)[†]**	**1.78 (1.34 - 2.36)**	A (vs. G)
TNF	rs1800629 (29)	A	UMNR	UMNR	**2.40 (1.98 - 2.90)**			**2.22 (1.85 - 2.66)[†]**	**3.25 (2.08 - 5.08)**	A (vs. G)
TNF	rs1800629 (29)	Non-A	UMNR	UMNR	0.97 (0.83 - 1.14)			1.00 (0.86 - 1.16)[†]	1.05 (0.71 - 1.55)	A (vs. G)
TNF	rs1800629 (29)	A, Non-A	S	UMNR	1.13 (0.95 - 1.35)			1.12 (0.91 - 1.37)[†]	1.45 (0.88 - 2.40)	A (vs. G)
TNF	rs1800629 (29)	A	S	UMNR	1.26 (0.69 - 2.30)			1.24 (0.85 - 1.82)[†]	1.66 (0.75 - 3.68)	A (vs. G)
TNF	rs1800629 (29)	Non-A	S	UMNR	1.06 (0.86 - 1.30)			1.07 (0.84 - 1.37)[†]	1.32 (0.69 - 2.53)	A (vs. G)
TNF	rs1800630 (30)	C, A	UMNR	UMNR	**0.76 (0.60 - 0.97)**		**0.74 (0.56 - 0.96)**			A (vs. C)
TNF	rs1800630 (30)	A	UMNR	UMNR	**0.56 (0.37 - 0.85)**		**0.50 (0.32 - 0.79)**			A (vs. C)

TNF	rs1800630 (30)	C	UMNR	UMNR	0.90 (0.67 - 1.22)	0.91 (0.65 - 1.28)	A (vs. C)

Results of most comprehensive genetic factors with strong significant associations (OR of ≤0.5 and ≥2.0) Legend: *Symbols*. *: if unspecified, it is an 'overall' estimate, including a mixed population of ethnicities, genders and/or smoking status. †: This heterozygous model comparison is WW vs. WM, rather than WM vs. MM. ‡: This estimate was inversed to fit the genetic model and comparison in relation to other estimates. $: rs1059046 (A>C), rs1136451 (A>G), rs4253527 (C>T), rs1130866 (C>T), rs2077079 (C>A), rs1051246 (A>G), rs2245121 (G>A), rs2255601 (G>A), rs3088308 (T>A), rs6413520 (A>G), rs721917 (T>C)and rs911887 (T>C).

Abbreviations. Outcomes: Genders: F = Female, M = Male, UMNR = Unclear/mixed/not reported; Smoking status: Non-S = Non-smokers, S = Smokers, UMNR = Unclear/mixed/not reported; Ethnicities: AA = African American, A = Asian, Af = African, Ar = Arabian, C = Caucasian, E = European, E-A = East Asian, I = Indian, S-A = South Asian, UMNR = Unclear/mixed/not reported. Estimates in bold face indicate significant associations.

Genetic models. M = Mutant; W = Wildtype. Allelic model: M vs. W; Recessive model = MM vs. MW+WW; Dominant model = MM+MW vs. WW; Heterozygous model = MW vs. WW; Homozygous model = MM vs. WW; Overdominant model = MW vs. MM+WW.

ORs displayed bold are significant ORs

Supplementary Table 4

Gene	Variant / rs number	Ethni-city	Smo-king *	Gen der *	Allelic	Recessiv e	Domina nt	Hetero-zygous	Homo-zygous	Over-domina nt	Other	Risk allele
ADAM19	rs113897301 [25]	UMNR	UMN R	UM NR	**1.16 (1.12 - 1.21)**							AT (vs. A)
ADAM19	rs1990950 [42]	C	UMN R	UM NR	**1.07 (1.04 - 1.10)**							G (vs. T)
ADAM33	rs3918396 (S1)[60]	E, A	UMN R	UM NR	1.05 (0.72 - 1.53)	0.97 (0.61 - 1.55)	**1.53 (1.06 - 2.21)**		**1.69 (1.23 - 2.32)**			C (vs. T)
ADAM33	rs3918396 (S1)[60]	E	UMN R	UM NR	0.81 (0.37 - 1.79)	0.80 (0.33 - 1.94)	0.69 (0.26 - 1.82)		0.67 (0.25 - 1.79)			C (vs. T)
ADAM33	rs3918396 (S1)[60]	A	UMN R	UM NR	**1.40 (1.20 - 1.64)**	1.17 (0.89 - 1.55)	**1.73 (1.21 - 2.46)**		**1.87 (1.34 - 2.60)**			C (vs. T)
ADAM33	rs511898 (F+1)[60]	E, A	UMN R	UM NR	**1.18 (1.02 - 1.38)**	**1.51 (1.12 - 2.05)**	1.11 (0.90 - 1.36)		**1.42 (1.06 - 1.90)**			T (vs. C)
ADAM33	rs511898 (F+1)[60]	E	UMN R	UM NR	1.20 (0.93 - 1.55)	1.39 (0.97 - 2.00)	1.20 (0.88 - 1.63)		1.48 (0.91 - 2.40)			T (vs. C)
ADAM33	rs511898 (F+1)[60]	A	UMN R	UM NR	1.14 (0.99 - 1.31)	1.68 (0.97 - 2.93)	1.01 (0.76 - 1.36)		1.32 (1.00 - 1.75)			T (vs. C)
ADAM33	rs597980 (ST+5)[60]	E	UMN R	UM NR	**1.25 (1.05 - 1.48)**	1.27 (0.94 - 1.71)	**1.38 (1.06 - 1.80)**		**1.51 (1.07 - 2.14)**			A (vs. G)
ADGRG6	rs9399401 [25]	UMNR	UMN R	UM NR	**1.15 (1.12 - 1.19)**							T (vs. C)
ADGRG6	rs7753012 [42]	C	UMN R	UM NR	**1.13 (1.10 - 1.16)**							T (vs. G)

ADGRG6	rs148274477 [42]	C	UMNR	UMNR	**1.21 (1.12 - 1.31)**					C (vs. T)
ADRB2	rs1042713 [37]	C, A, Af	UMNR	UMNR				1.01 (0.81 - 1.26)	0.97 (0.76 - 1.22)	G (vs. A)
ADRB2	rs1042713 [46]	C, A	UMNR	UMNR	1.07 (0.86 - 1.34)	1.09 (0.76 - 1.55)	1.10 (0.81 - 1.47)	1.06 (0.77 - 1.47)†‡	1.15 (0.72 - 1.82)‡	A (vs. G)
ADRB2	rs1042713 [46]	C	UMNR	UMNR	1.02 (0.72 - 1.44)	1.03 (0.63 - 1.67)	1.00 (0.61 - 1.61)	0.95 (0.71 - 1.28)†‡	1.04 (0.50 - 2.13)‡	A (vs. G)
ADRB2	rs1042713 [46]	A	UMNR	UMNR	1.14 (0.86 - 1.52)	1.13 (0.65 - 1.96)	1.30 (0.91 - 1.85)	1.09 (0.62 - 1.89)†‡	1.39 (0.93 - 2.13)‡	A (vs. G)
ADRB2	rs1042713 [46]	C, A	S	UMNR	1.04 (0.84 - 1.30)	1.10 (0.81 - 1.50)	0.99 (0.73 - 1.33)	1.09 (0.86 - 1.37)†‡	1.05 (0.67 - 1.67)‡	A (vs. G)
ADRB2	rs1042713 [46]	C	S	UMNR	0.88 (0.66 - 1.17)	0.83 (0.62 - 1.11)	0.83 (0.54 - 1.28)	0.88 (0.65 - 1.19)†‡	0.76 (0.43 - 1.37)‡	A (vs. G)
ADRB2	rs1042713 [46]	A	S	UMNR	**1.27 (1.03 - 1.57)**	**1.45 (1.04 - 2.01)**	1.27 (0.88 - 1.85)	**1.43 (1.01 - 2.00)†‡**	**1.56 (1.01 - 2.44)‡**	A (vs. G)
AGER	rs2070600 [42]	C	UMNR	UMNR	**1.20 (1.14 - 1.27)**					C (vs. T)
AGPHD1	rs8042849 [20]	C, AA	S	M	**0.76 (0.66 - 0.85)**					T (vs. C)
AGPHD1	rs8042849 [20]	C, AA	S	F	**0.84 (0.74 - 0.94)**					T (vs. C)
AGPHD1	rs9788721 [20]	C, AA	S	M	**0.77 (0.68 - 0.86)**					T (vs. C)
AGPHD1	rs9788721 [20]	C, AA	S	F	**0.84 (0.75 - 0.94)**					T (vs. C)

AKD1	rs10499052 [24]	UMNR	UMNR	UMNR	**1.16 (1.09 - 1.23)**	A (vs. G)
ARL15	rs2441026 [42]	C	UMNR	UMNR	**1.04 (1.00 - 1.07)**	C (vs. T)
ARMC2	rs2768551 [42]	C	UMNR	UMNR	**1.08 (1.04 - 1.11)**	A (vs. G)
ARMC2	rs2806356 [25]	UMNR	UMNR	UMNR	**1.12 (1.08 - 1.16)**	C (vs. T/G)
ASTN2	rs803923 [42]	C	UMNR	UMNR	**1.06 (1.03 - 1.08)**	A (vs. G)
C5orf56	rs7713065 [42]	C	UMNR	UMNR	**1.05 (1.01 - 1.10)**	A (vs. C)
CACNA2D3/ WNT5A	rs1458979 [42]	C	UMNR	UMNR	**1.06 (1.02 - 1.10)**	G (vs. A)
CCDC101	rs17707300 [25]	UMNR	UMNR	UMNR	**1.10 (1.06 - 1.13)**	C (vs. T)
CDC123	rs7090277 [42]	C	UMNR	UMNR	**1.07 (1.05 - 1.10)**	T (vs. A)
CELSR1	rs56344079 [20]	C, AA	S	M	1.08 (0.96 - 1.19)	C (vs. G/T)
CELSR1	rs56344079 [20]	C, AA	S	F	**0.72 (0.60 - 0.85)**	C (vs. G/T)
CELSR1	rs7286446 [20]	C, AA	S	M	**0.87 (0.76 - 0.98)**	T (vs. C)
CELSR1	rs7286446 [20]	C, AA	S	F	**1.31 (1.18 - 1.43)**	T (vs. C)
CELSR1	rs9615358 [20]	C, AA	S	M	0.90 (0.79 - 1.01)	A (vs. G)
CELSR1	rs9615358 [20]	C, AA	S	F	**1.37 (1.25 - 1.49)**	A (vs. G)
CELSR1	rs9615973 [20]	C, AA	S	M	1.06 (0.94 - 1.17)	C (vs. G)

CELSR1	rs9615973 [20]	C, AA	S	F	**0.73 (0.61 - 0.86)**			C (vs. G)
CELSR1	rs9615981 [20]	C, AA	S	M	**0.87 (0.76 - 0.98)**			T (vs. G/A)
CELSR1	rs9615981 [20]	C, AA	S	F	**1.31 (1.18 - 1.43)**			T (vs. G/A)
CELSR1	rs9615982 [20]	C, AA	S	M	**0.86 (0.71 - 0.98)**			T (vs. G/C)
CELSR1	rs9615982 [20]	C, AA	S	F	**1.31 (1.18 - 1.43)**			T (vs. G/C)
CFDP1	rs7186831 [25]	UMNR	UMNR	UMNR	**1.12 (1.08 - 1.16)**			A (vs. G)
CFDP1	rs3743609 [42]	C	UMNR	UMNR	**1.06 (1.03 - 1.09)**			C (vs. G)
CHRNA3	rs1051730 [20]	C, AA	S	M	**1.29 (1.20 - 1.39)**			A (vs. G)
CHRNA3	rs1051730 [20]	C, AA	S	F	**1.19 (1.09 - 1.30)**			A (vs. G)
CHRNA3	rs1051730 [11]	A, C, African, unspecified	UMNR	UMNR	**1.14 (1.10 - 1.18)**	**1.13 (1.07 - 1.19)** Multivariate adjusted: **1.11 (1.01 - 1.20)**	**1.30 (1.20 - 1.41)**	A (vs. G)
CHRNA3	rs1051730 [11]	A	UMNR	UMNR	1.23 (0.91 - 1.67)			A (vs. G)
CHRNA3	rs1051730 [11]	Non-A	UMNR	UMNR	**1.14 (1.10 - 1.18)**			A (vs. G)
CHRNA3	rs114205691 [20]	C, AA	S	M	**1.33 (1.23 - 1.42)**			T (vs. C/A)
CHRNA3	rs114205691 [20]	C, AA	S	F	**1.21 (1.11 - 1.31)**			T (vs. C/A)
CHRNA3	rs12914385 [20]	C, AA	S	M	**1.32 (1.22 - 1.41)**			T (vs. C/A)

CHRNA3	rs12914385 [20]	C, AA	S	F	**1.23 (1.13 - 1.32)**	T (vs. C/A)
CHRNA3	rs12914385, Moderate-to-severe COPD [10]	C, AA	UMNR	UMNR	**1.28 (1.20 - 1.36)**	T (vs. C/A)
CHRNA3	rs12914385, Severe COPD [10]	C, AA	UMNR	UMNR	**1.39 (1.29 - 1.51)**	T (vs. C/A)
CHRNA3	rs138544659 [20]	C, AA	S	M	**0.74 (0.63 - 0.84)**	T (vs. G)
CHRNA3	rs138544659 [20]	C, AA	S	F	**0.82 (0.71 - 0.93)**	T (vs. G)
CHRNA3	rs141518190 [20]	C, AA	S	M	**0.74 (0.64 - 0.85)**	A (vs. G)
CHRNA3	rs141518190 [20]	C, AA	S	F	**0.83 (0.72 - 0.94)**	A (vs. G)
CHRNA3	rs146009840 [20]	C, AA	S	M	**0.77 (0.67 - 0.87)**	A (vs. T)
CHRNA3	rs146009840 [20]	C, AA	S	F	**0.83 (0.72 - 0.93)**	A (vs. T)
CHRNA3	rs147144681 [20]	C, AA	S	M	**1.35 (1.25 - 1.45)**	T (vs. C)
CHRNA3	rs147144681 [20]	C, AA	S	F	**1.20 (1.10 - 1.31)**	T (vs. C)
CHRNA3	rs147499554 [20]	C, AA	S	M	**1.35 (1.24 - 1.45)**	T (vs. C/G)
CHRNA3	rs147499554 [20]	C, AA	S	F	**1.21 (1.10 - 1.32)**	T (vs. C/G)
CHRNA3	rs4887067 [20]	C, AA	S	M	**1.31 (1.21 - 1.41)**	A (vs. G)
CHRNA3	rs4887067 [20]	C, AA	S	F	**1.20 (1.09 - 1.30)**	A (vs. G)

CHRNA3	rs55676755 [20]	C, AA	S	M	**0.75 (0.66 - 0.85)**			C (vs. G)
CHRNA3	rs55676755 [20]	C, AA	S	F	**0.82 (0.72 - 0.92)**			C (vs. G)
CHRNA3	rs56077333 [20]	C, AA	S	M	**1.35 (1.26 - 1.45)**			A (vs. C/T)
CHRNA3	rs56077333 [20]	C, AA	S	F	**1.22 (1.12 - 1.33)**			A (vs. C/T)
CHRNA3	rs6495309 [11]	A	UMNR	UMNR	**1.26 (1.09 - 1.45)**	1.12 (0.85 - 1.49)	**1.53 (1.14 - 2.06)** Multivariate adjusted: **1.31 (1.08 - 1.54)**	C (vs. T/A)
CHRNA3	rs8192482 [20]	C, AA	S	M	**1.31 (1.21 - 1.41)**			T (vs. C)
CHRNA3	rs8192482 [20]	C, AA	S	F	**1.19 (1.09 - 1.30)**			T (vs. C)
CHRNA5	rs11633958 [20]	C, AA	S	M	**1.30 (1.20 - 1.40)**			T (vs. C/A)
CHRNA5	rs11633958 [20]	C, AA	S	F	**1.19 (1.08 - 1.29)**			T (vs. C/A)
CHRNA5	rs140330585 [20]	C, AA	S	M	**1.30 (1.20 - 1.40)**			A (vs. G)
CHRNA5	rs140330585 [20]	C, AA	S	F	**1.20 (1.09 - 1.30)**			A (vs. G)
CHRNA5	rs16969968 MMA [11] [24]	UMNR	UMNR	UMNR	**1.30 (1.24 – 1.36); I^2=0.0%; n=2**			A (vs. G)
CHRNA5	rs16969968 [20]	C, AA	S	M	**1.30 (1.20 - 1.40)**			A (vs. G)

CHRNA5	rs16969968 [20]	C, AA	S	F	**1.19 (1.09 - 1.30)**	A (vs. G)
CHRNA5	rs17486195 [20]	C, AA	S	M	**0.78 (0.68 - 0.87)**	A (vs. G/T)
CHRNA5	rs17486195 [20]	C, AA	S	F	**0.84 (0.73 - 0.94)**	A (vs. G/T)
CHRNA5	rs17486278 [25]	UMNR	UMNR	UMNR	**1.18 (1.15 - 1.22)**	C (vs. A)
CHRNA5	rs17486278 [20]	C, AA	S	M	**0.76 (0.67 - 0.85)**	A (vs. C)
CHRNA5	rs17486278 [20]	C, AA	S	F	**0.82 (0.73 - 0.92)**	A (vs. C)
CHRNA5	rs190065944 [20]	C, AA	S	M	**1.39 (1.26 - 1.52)**	A (vs. G)
CHRNA5	rs190065944 [20]	C, AA	S	F	**1.26 (1.12 - 1.39)**	A (vs. G)
CHRNA5	rs2036527 [20]	C, AA	S	M	**1.28 (1.19 - 1.37)**	A (vs. G)
CHRNA5	rs2036527 [20]	C, AA	S	F	**1.18 (1.09 - 1.28)**	A (vs. G)
CHRNA5	rs55853698 [20]	C, AA	S	M	**0.77 (0.67 - 0.87)**	T (vs. G)
CHRNA5	rs55853698 [20]	C, AA	S	F	**0.84 (0.74 - 0.94)**	T (vs. G)
CHRNA5	rs56390833 [20]	C, AA	S	M	**1.30 (1.20 - 1.40)**	A (vs. C/T)
CHRNA5	rs56390833 [20]	C, AA	S	F	**1.20 (1.09 - 1.30)**	A (vs. C/T)
CHRNA5	rs7172118 [20]	C, AA	S	M	**1.30 (1.20 - 1.40)**	A (vs. C/T)
CHRNA5	rs7172118 [20]	C, AA	S	F	**1.20 (1.10 - 1.30)**	A (vs. C/T)

CHRNA5	rs7180002 [20]	C, AA	S	M	**0.77 (0.67 - 0.87)**	A (vs. T)
CHRNA5	rs7180002 [20]	C, AA	S	F	**0.84 (0.73 - 0.94)**	A (vs. T)
CHRNA5	rs72740955 [20]	C, AA	S	M	**1.29 (1.19 - 1.39)**	T (vs. C)
CHRNA5	rs72740955 [20]	C, AA	S	F	**1.19 (1.09 - 1.29)**	T (vs. C)
CHRNA5	rs72740964 [20]	C, AA	S	M	**1.30 (1.20 - 1.40)**	A (vs. G)
CHRNA5	rs72740964 [20]	C, AA	S	F	**1.20 (1.09 - 1.30)**	A (vs. G)
CHRNA5	rs951266 [20]	C, AA	S	M	**1.30 (1.20 - 1.40)**	A (vs. G)
CHRNA5	rs951266 [20]	C, AA	S	F	**1.19 (1.09 - 1.30)**	A (vs. G)
CHRNB4	rs55988292 [20]	C, AA	S	M	**0.77 (0.67 - 0.87)**	A (vs. G)
CHRNB4	rs55988292 [20]	C, AA	S	F	**0.88 (0.78 - 0.98)**	A (vs. G)
CHRNB4	rs72743158 [20]	C, AA	S	M	**0.77 (0.67 - 0.87)**	T (vs. C)
CHRNB4	rs72743158 [20]	C, AA	S	F	**0.85 (0.74 - 0.96)**	T (vs. C)
CHRNB4	rs17487223 [20]	C, AA	S	M	**1.31 (1.21 - 1.41)**	T (vs. C)
CHRNB4	rs17487223 [20]	C, AA	S	F	**1.20 (1.10 - 1.30)**	T (vs. C)
CRAMP1L	rs61746451[24]	UMNR	UMNR	UMNR	**0.57 (0.29 - 0.84)**	T (vs. C)
CYFIP2	rs10515750 [42]	C	UMNR	UMNR	**1.13 (1.05 - 1.22)**	T (vs. C)

CYP1A1	rs4646903 [43]	C, A	UMNR	UMNR	1.21 (0.91 - 1.60)	**1.57 (1.09 - 2.26)**	1.14 (0.77 - 1.70)	1.22 (0.77 - 1.94)	**1.73 (1.18 - 2.55)**	C (vs. T)
CYP1A1	rs4646903 [43]	A	UMNR	UMNR	1.29 (0.94 - 1.78)	**1.70 (1.06 - 2.71)**	1.07 (0.59 - 1.92)	1.22 (0.51 - 2.89)	**1.84 (1.11 - 3.06)**	C (vs. T)
CYP1A1	rs4646903 [43]	C	UMNR	UMNR	1.13 (0.69 - 1.86)	1.39 (0.78 - 2.48)	1.23 (0.63 - 2.40)	1.24 (0.61 - 2.51)	1.60 (0.89 - 2.88)	C (vs. T)
CYP2A6	rs12459249 [25]	UMNR	UMNR	UMNR	**1.10 (1.06 - 1.14)**					C (vs. T)
DSP	rs2076295 [25]	UMNR	UMNR	UMNR	**1.09 (1.06 - 1.12)**					T (vs. G)
EEFSEC	rs2811416 [20]	C, AA	S	M	**0.84 (0.69 - 0.99)**					T (vs. C/G)
EEFSEC	rs2811416 [20]	C, AA	S	F	**0.66 (0.50 - 0.82)**					T (vs. C/G)
EEFSEC	rs2955083 [25]	UMNR	UMNR	UMNR	**1.18 (1.13 - 1.24)**					A (vs. T)
EEFSEC	rs2811415 [42]	C	UMNR	UMNR	**1.05 (1.00 - 1.11)**					G (vs. A)
EPB41L4A-AS1	rs66669542 [20]	C, AA	S	M	**0.76 (0.65 - 0.87)**					A (vs. T)
EPB41L4A-AS1	rs66669542 [20]	C, AA	S	F	0.94 (0.82 - 1.05)					A (vs. T)
EPHX1	rs1051740 [31]	C, A	UMNR	UMNR				1.12 (0.96 - 1.30)	**1.33 (1.06 - 1.69)**	C (vs. T)
EPHX1	rs1051740 [31]	A	UMNR	UMNR				1.07 (0.69 - 1.65)	1.07 (0.76 - 1.52)	C (vs. T)
EPHX1	rs1051740 [31]	C	UMNR	UMNR				1.08 (0.96 - 1.22)	**1.61 (1.12 - 2.31)**	C (vs. T)
FAM13A	rs10021465 [20]	C, AA	S	M	**0.87 (0.78 - 0.96)**					A (vs. G)

FAM13A	rs10021465 [20]	C, AA	S	F	**0.76 (0.66 - 0.86)**	A (vs. G)
FAM13A	rs1812329 [20]	C, AA	S	M	**1.19 (1.11 - 1.28)**	A (vs. G)
FAM13A	rs1812329 [20]	C, AA	S	F	**1.35 (1.25 - 1.45)**	A (vs. G)
FAM13A	rs1964516 [9]	C	UMNR	UMNR	**0.73 (0.66 - 0.81)**	C (vs. T)
FAM13A	rs2013701 [20]	C, AA	S	M	**1.18 (1.10 - 1.27)**	T (vs. G)
FAM13A	rs2013701 [20]	C, AA	S	F	**1.30 (1.20 - 1.39)**	T (vs. G)
FAM13A	rs2045517 [20]	C, AA	S	M	**1.18 (1.10 - 1.27)**	T (vs. C)
FAM13A	rs2045517 [20]	C, AA	S	F	**1.35 (1.25 - 1.45)**	T (vs. C)
FAM13A	rs2045517 [42]	C	UMNR	UMNR	**1.10 (1.08 - 1.13)**	T (vs. C)
FAM13A	rs28455964 [20]	C, AA	S	M	**0.85 (0.76 - 0.94)**	T (vs. G)
FAM13A	rs28455964 [20]	C, AA	S	F	**0.78 (0.68 - 0.87)**	T (vs. G)
FAM13A	rs2869966 [20]	C, AA	S	M	**1.19 (1.11 - 1.28)**	T (vs. C)
FAM13A	rs2869966 [20]	C, AA	S	F	**1.35 (1.26 - 1.45)**	T (vs. C)
FAM13A	rs2869967 [20]	C, AA	S	M	**0.84 (0.75 - 0.93)**	T (vs. C)
FAM13A	rs2869967 [20]	C, AA	S	F	**0.74 (0.64 - 0.84)**	T (vs. C)
FAM13A	rs2904259 [20]	C, AA	S	M	**0.85 (0.76 - 0.93)**	T (vs. C/A)

FAM13A	rs2904259 [20]	C, AA	S	F	**0.78 (0.68 - 0.87)**	T (vs. C/A)
FAM13A	rs3846287 [20]	C, AA	S	M	**1.17 (1.08 - 1.26)**	T (vs. C/A/G)
FAM13A	rs3846287 [20]	C, AA	S	F	**1.28 (1.18 - 1.37)**	T (vs. C/A/G)
FAM13A	rs3857043 [20]	C, AA	S	M	**0.85 (0.76 - 0.94)**	T (vs. C)
FAM13A	rs3857043 [20]	C, AA	S	F	**0.78 (0.68 - 0.88)**	T (vs. C)
FAM13A	rs4416442 [20]	C, AA	S	M	**0.83 (0.74 - 0.91)**	T (vs. C)
FAM13A	rs4416442 [20]	C, AA	S	F	**0.74 (0.65 - 0.84)**	T (vs. C)
FAM13A	rs4416442, Moderate-to-severe COPD [10]	C, AA	UMNR	UMNR	**1.28 (1.20 - 1.36)**	C (vs. T)
FAM13A	rs4416442, Severe COPD [10]	C, AA	UMNR	UMNR	**1.36 (1.26 - 1.47)**	C (vs. T)
FAM13A	rs4693980 [20]	C, AA	S	M	**1.17 (1.08 - 1.25)**	A (vs. G)
FAM13A	rs4693980 [20]	C, AA	S	F	**1.29 (1.20 - 1.39)**	A (vs. G)
FAM13A	rs6830970 [20]	C, AA	S	M	**0.86 (0.77 - 0.94)**	A (vs. G)
FAM13A	rs6830970 [20]	C, AA	S	F	**0.78 (0.68 - 0.88)**	A (vs. G)
FAM13A	rs6837671 [20]	C, AA	S	M	**0.83 (0.74 - 0.92)**	A (vs. G)
FAM13A	rs6837671 [20]	C, AA	S	F	**0.74 (0.64 - 0.84)**	A (vs. G)

FAM13A	rs6837671 [25]	UMNR	UMNR	UMNR	**1.12 (1.09-1.15)**	G (vs. A)
FAM13A	rs76273989 [20]	C, AA	S	M	**1.62 (1.43 - 1.81)**	A (vs. C)
FAM13A	rs76273989 [20]	C, AA	S	F	1.20 (0.99 - 1.40)	A (vs. C)
FAM13A	rs7671167 [9]	C	UMNR	UMNR	**0.73 (0.66 - 0.81)**	C (vs. T)
FAM13A	rs7671167 [20]	C, AA	S	M	**1.17 (1.09 - 1.26)**	T (vs. C)
FAM13A	rs7671167 [20]	C, AA	S	F	**1.30 (1.20 - 1.39)**	T (vs. C)
FAM13A	rs7671261 [20]	C, AA	S	M	**1.17 (1.08 - 1.26)**	A (vs. G)
FAM13A	rs7671261 [20]	C, AA	S	F	**1.28 (1.18 - 1.37)**	A (vs. G)
FAM13A	rs7674369 [20]	C, AA	S	M	**1.20 (1.11 - 1.29)**	A (vs. G/T)
FAM13A	rs7674369 [20]	C, AA	S	F	**1.35 (1.25 - 1.44)**	A (vs. G/T)
FAM13A	rs7682317 [20]	C, AA	S	M	**1.21 (1.12 - 1.30)**	T (vs. C)
FAM13A	rs7682317 [20]	C, AA	S	F	**1.35 (1.25 - 1.44)**	T (vs. C)
FAM13A	rs7682431 [20]	C, AA	S	M	**1.16 (1.07 - 1.25)**	C (vs. G)
FAM13A	rs7682431 [20]	C, AA	S	F	**1.32 (1.22 - 1.42)**	C (vs. G)
FAM13A	rs13110699 [42]	C	UMNR	UMNR	**1.15 (1.09 - 1.21)**	G (vs. T)
FAM208B	rs41290259 [24]	UMNR	UMNR	UMNR	**1.79 (1.49-2.08)**	A (vs. G)

FAR2	rs7294481 [20]	C, AA	S	M	**0.83 (0.74 - 0.92)**	T (vs. C)
FAR2	rs7294481 [20]	C, AA	S	F	**1.15 (1.06 - 1.25)**	T (vs. C)
FOXO1	rs75700692 [20]	C, AA	S	M	1.02 (0.87 - 1.17)	A (vs. C)
FOXO1	rs75700692 [20]	C, AA	S	F	**0.65 (0.49 - 0.82)**	A (vs. C)
FOXO1	rs78372177 [20]	C, AA	S	M	1.00 (0.85 - 1.16)	A (vs. G)
FOXO1	rs78372177 [20]	C, AA	S	F	**0.63 (0.45 - 0.80)**	A (vs. G)
GEMIN4	rs11652959 [20]	C, AA	S	M	1.07 (0.94 - 1.20)	T (vs. C/A/G)
GEMIN4	rs11652959 [20]	C, AA	S	F	**1.43 (1.30 - 1.57)**	T (vs. C/A/G)
GLIS3	rs7872188 [42]	C	UMNR	UMNR	**1.06 (1.02 - 1.10)**	T (vs. C)
GSTCD	rs11727735 [25]	UMNR	UMNR	UMNR	**1.26 (1.18 - 1.33)**	A (vs. G)
GSTCD	rs10516526 [42]	C	UMNR	UMNR	**1.12 (1.07 - 1.18)**	A (vs. G)
GYPA	rs13105210 [20]	C, AA	S	M	**0.84 (0.74 - 0.94)**	T (vs. C)
GYPA	rs13105210 [20]	C, AA	S	F	**0.75 (0.64 - 0.85)**	T (vs. C)
GYPA	rs4835177 [20]	C, AA	S	M	**1.18 (1.07 - 1.28)**	A (vs. G)
GYPA	rs4835177 [20]	C, AA	S	F	**1.32 (1.21 - 1.43)**	A (vs. G)

GSTT1	Null/Wt [15]	C, A, Af	UMNR	UMNR						**1.28 (1.09 – 1.50)**	Null (vs. Wt)
GSTM1 & GSTT1 combined	Null/Wt [15]	C, A, Af	UMNR	UMNR						**1.42 (1.21 – 1.66)**	Null (vs. Wt)
HDAC4	rs12477314 [42]	C	UMNR	UMNR	**1.09 (1.06 - 1.12)**						T (vs. C)
HHIP	rs13118928 [9]	C	UMNR	UMNR	**0.76 (0.65 - 0.89)**						G (vs. A)
HHIP	rs13141641 [25]	UMNR	UMNR	UMNR	**1.22 (1.19 - 1.26)**						T (vs. C)
HHIP	rs1828591 [8]	C, AA	UMNR	UMNR	**0.77 (0.70 – 0.86)**						G (vs. A)
HLA-DQB1/HLA-DQA2	rs34864796 [42]	C	UMNR	UMNR	**1.12 (1.07 - 1.16)**						A (vs. G)
HLA-DQB1/HLA-DQA2	rs114229351 [42]	C	UMNR	UMNR	**1.07 (1.00 - 1.13)**						C (vs. T)
HTR4	rs7715901 [42]	C	UMNR	UMNR	**1.10 (1.08 - 1.13)**						A (vs. G)
HTR4	rs7733088 [25]	UMNR	UMNR	UMNR	**1.18 (1.14 - 1.21)**						G (vs. A/C)
HYKK	rs8034191 [11]	C, A, AA, Unspecified	UMNR	UMNR	**1.29 (1.18 - 1.41)**			**1.37 (1.20 - 1.56)**	**1.56 (1.27 - 1.92)**		C (vs. T)
IL1B	rs1143627 [51]	E-A, S-A, Ar	UMNR	UMNR		0.80 (0.52 - 1.21)	1.25 (0.79 - 1.96)		1.00 (0.52 - 1.95)		G (vs. A)
IL1B	rs1143627 [51]	E-A	UMNR	UMNR		0.87 (0.64 - 1.18)	**1.55 (1.14 - 2.11)**		1.27 (0.87 - 1.87)		G (vs. A)

IL1B	rs16944 [44]	C, A	UMNR	UMNR			0.89 (0.78 - 1.01)			A (vs. G)
IL1B	rs16944 [44]	A	UMNR	UMNR			**0.73 (0.60 - 0.88)**			A (vs. G)
IL1B	rs16944 [44]	C	UMNR	UMNR			1.07 (0.89 - 1.29)			A (vs. G)
IL6	rs1800795 [50]	C	UMNR	UMNR	**1.16 (1.03 - 1.30)**	1.21 (0.96 - 1.53)	**1.21 (1.02 - 1.43)**	1.18 (0.99 - 1.40)	**1.32 (1.03 - 1.70)**	C (vs. G)
IL27	rs181206 [24]	UMNR	UMNR	UMNR	**0.85 (0.78 - 0.92)**					A (vs. G)
IREB2	rs1062980 [8]	C, AA	UMNR	UMNR	**0.80 (0.69 – 0.87)**					C (vs. T)
IREB2	rs11858836 [9]	C	UMNR	UMNR	**1.29 (1.13 - 1.47)**					A (vs. G/T)
IREB2	rs13180 [9]	C	UMNR	UMNR	**0.78 (0.70 - 0.86)**					C (vs. T)
IREB2	rs2568494 [20]	C, AA	S	M	**0.79 (0.74 - 0.85)‡**					G (vs. A)
IREB2	rs2568494 [20]	C, AA	S	F	**0.88 (0.81 - 0.96)‡**					G (vs. A)
IREB2	rs2568494 [16]	C, I, A	UMNR	UMNR	0.95 (0.77 - 1.18)	0.94 (0.75 - 1.18)	**0.72 (0.57-0.92)‡**	0.91 (0.79 - 1.04)	0.88 (0.50 - 1.56)	G (vs. A)
IREB2 (MMA)	rs2568494 [16 20]	C, I, A, AA	UMNR	UMNR	**0.85 (0.76 - 0.93); I^2=60.3%; n=3**					G (vs. A)
IREB2	rs2656052 [20]	C, AA	S	M	**0.79 (0.70 - 0.88)**					A (vs. C)

IREB2	rs2656052 [20]	C, AA	S	F	**0.88 (0.79 - 0.98)**	A (vs. C)
IREB2	rs2656065 [20]	C, AA	S	M	**1.26 (1.16 - 1.35)**	A (vs. G)
IREB2	rs2656065 [20]	C, AA	S	F	**1.13 (1.04 - 1.23)**	A (vs. G)
IREB2	rs2938670 [20]	C, AA	S	M	**0.79 (0.70 - 0.88)**	T (vs. G)
IREB2	rs2938670 [20]	C, AA	S	F	**0.88 (0.79 - 0.98)**	T (vs. G)
ITGA1	rs1551943 [42]	C	UMNR	UMNR	**1.08 (1.03 - 1.12)**	A (vs. G)
LINC00310/KCNE2	rs2834440 [42]	C	UMNR	UMNR	**1.05 (1.03 - 1.08)**	G (vs. A)
LOC105372926	rs4328080 [42]	C	UMNR	UMNR	**1.04 (1.02 - 1.07)**	G (vs. A)
LOC105377462	rs138641402 [42]	C	UMNR	UMNR	**1.17 (1.14 - 1.20)**	A (vs. T)
LOC107984437	rs10850377 [42]	C	UMNR	UMNR	**1.03 (1.00 - 1.05)**	G (vs. A)
LOC389602/LOC285889	rs12698403 [42]	C	UMNR	UMNR	**1.05 (1.01 - 1.09)**	A (vs. G)
LOC105377462	rs720485 [8]	C, AA	UMNR	UMNR	**0.78 (0.64 – 0.81)**	C (vs. A/T)
LRMDA	rs2637254 [42]	C	UMNR	UMNR	**1.05 (1.02 - 1.07)**	A (vs. G)
LRP1	rs11172113 [42]	C	UMNR	UMNR	**1.04 (1.01 - 1.07)**	T (vs. C)
KBTBD12	rs17282209 [20]	C, AA	S	M	**1.20 (1.05 - 1.36)**	T (vs. C)
KBTBD12	rs17282209 [20]	C, AA	S	F	**1.51 (1.35 - 1.67)**	T (vs. C)

KCNQ5	rs141651520 [42]	C	UMNR	UMNR	**1.05 (1.00 - 1.10)**					TTCTAT (vs. Del)
KCNS3/RDH14	rs62126408 [42]	C	UMNR	UMNR	**1.07 (1.04 - 1.11)**					T (vs. C)
MECOM/LOC100507661	rs56341938 [42]	C	UMNR	UMNR	**1.05 (1.01 - 1.10)**					A (vs. G)
MFAP2	rs2284746 [42]	C	UMNR	UMNR	**1.05 (1.02 - 1.07)**					G (vs. C)
MICAL1	rs59056467 [24]	UMNR	UMNR	UMNR	**1.14 (1.07-1.20)**					T (vs. C)
MMP3	rs679620 [24]	UMNR	UMNR	UMNR	**1.13 (1.07-1.19)**					T (vs. C)
MMP9	rs17576 [7]	C, A	UMNR	UMNR	**1.35 (1.00 - 1.82)**	1.20 (0.62 - 2.30)	**1.46 (1.02 - 2.08)**			A (vs. G)
MMP9	rs17576 [7]	A	UMNR	UMNR	1.44 (0.96 - 2.17)	0.78 (0.31 - 1.99)	**1.66 (1.01 - 2.71)**			A (vs. G)
MMP9	rs17576 [7]	C	UMNR	UMNR	1.13 (0.88 - 1.45)	2.23 (0.93 - 5.33)	1.07 (0.81 - 1.41)			A (vs. G)
MMP9	rs3918242 [27]	C, A	UMNR	UMNR	**1.47 (1.21 - 1.79)**	1.33 (0.82 - 2.15)	**1.36 (1.16 - 1.61)**	1.00 (0.60 - 1.66)[†]	1.56 (0.95 - 2.56)	T (vs. C)
MMP12	rs626750, severe COPD [10]	C, AA	UMNR	UMNR	**1.36 (1.23 - 1.51)**					G (vs.A)
MMP15	rs12447804 [42]	C	UMNR	UMNR	**1.04 (1.01 - 1.08)**					T (vs. C)
MN1	rs134041 [42]	C	UMNR	UMNR	**1.06 (1.03 - 1.08)**					T (vs. C)
MN1	rs2283847 [42]	C	UMNR	UMNR	**1.05 (1.01 - 1.09)**					T (vs.C)

MTCL1	rs647097 [25]	UMNR	UMNR	UMNR	**1.10 (1.06 - 1.13)**	C (vs. T)
NCR3/AIF1	rs2857595 [42]	C	UMNR	UMNR	**1.08 (1.05 - 1.11)**	A (vs. G)
NPNT	rs34712979 [42]	C	UMNR	UMNR	**1.12 (1.06 - 1.18)**	A (vs. G)
PABPC4	rs17513135 [42]	C	UMNR	UMNR	**1.06 (1.01 - 1.11)**	T (vs. C)
PID1	rs16825267 [25]	UMNR	UMNR	UMNR	**1.19 (1.12 - 1.25)**	C (vs. G/A)
PRDM11	rs2863171 [42]	C	UMNR	UMNR	**1.05 (1.02 - 1.09)**	A (vs. C)
PSMA4	rs58365910 [20]	C, AA	S	M	**0.78 (0.68 - 0.87)**	T (vs. C)
PSMA4	rs58365910 [20]	C, AA	S	F	**0.85 (0.75 - 0.94)**	T (vs. C)
PTCH1	rs16909859 [42]	C	UMNR	UMNR	**1.05 (1.01 - 1.10)**	A (vs. G)
RAB4B	rs2604894 [9]	C	UMNR	UMNR	**0.74 (0.65 - 0.84)**	A (vs. G)
RAB4B	rs7937 [9]	C	UMNR	UMNR	**0.73 (0.63 - 0.83)**	C (vs. T)
RARB	rs1529672 [42]	C	UMNR	UMNR	**1.07 (1.03 - 1.10)**	C (vs. A)
RIN3	rs1075472 [20]	C, AA	S	M	**1.13 (1.01 - 1.25)**	A (vs. G/T)
RIN3	rs1075472 [20]	C, AA	S	F	**1.39 (1.26 - 1.52)**	A (vs. G/T)
RIN3	rs72699855 [20]	C, AA	S	M	0.90 (0.78 - 1.02)	C (vs. G/A)
RIN3	rs72699855 [20]	C, AA	S	F	**0.72 (0.59 - 0.85)**	C (vs. G/A)

RIN3	rs754388 [25]	UMNR	UMNR	UMNR	**1.15 (1.11 - 1.20)**	C (vs. G/T)
RIN3	rs754388 [20]	C, AA	S	M	**1.17 (1.05 - 1.28)**	C (vs. G/T)
RIN3	rs754388 [20]	C, AA	S	F	**1.41 (1.28 - 1.54)**	C (vs. G/T)
RIN3	rs117068593 [42]	C	UMNR	UMNR	**1.08 (1.04 - 1.11)**	C (vs. T)
RSRC1	rs1595029 [42]	C	UMNR	UMNR	**1.04 (1.01 - 1.06)**	C (vs. A)
SERPINA1	rs28929474 [24]	UMNR	UMNR	UMNR	**1.57 (1.33 - 1.80)**	T (vs. C/G)
SERPINA1	PI MS [14]	C	UMNR	UMNR	**1.19 (1.02 - 1.38)**	MS (vs. MM)
SERPINA1	PI MS [14]	C	S	UMNR	1.02 (0.81 - 1.28)	MS (vs. MM)
SFTPD	rs721917 [25]	UMNR	UMNR	UMNR	**1.08 (1.05 - 1.11)**	G (vs. A)
SLC22A11	rs141159367 [36]	UMNR	UMNR	UMNR	**1.87 (1.49-2.26)**	T (vs. C/A)
SNRPF	rs12820313 [42]	C	UMNR	UMNR	**1.06 (1.03 - 1.10)**	C (vs. T)
SPATA9	rs153916 [42]	C	UMNR	UMNR	**1.04 (1.02 - 1.07)**	T (vs. C)
SPHKAP/PID1	rs10498230 [42]	C	UMNR	UMNR	**1.12 (1.07 - 1.18)**	C (vs. T)
TEKT5	rs12149828 [42]	C	UMNR	UMNR	**1.06 (1.02 - 1.09)**	A (vs. G)
TESMIN	rs146043252 [36]	UMNR	UMNR	UMNR	**1.66 (1.17 – 2.15)**	G (vs. A)
TET2	rs2047409 [25]	UMNR	UMNR	UMNR	**1.12 (1.08 - 1.15)**	A (vs. G)

TGFB1	rs1800470 [34]	C, A	UMNR	UMNR	0.87 (0.73 - 1.03)	0.94 (0.51 - 1.74)	0.81 (0.59 - 1.12)	0.74 (0.40 - 1.37)	C (vs. T/G)
TGFB1	rs1800470 [34]	C	UMNR	UMNR	**0.79 (0.64 - 0.99)**	0.61 (0.26 - 1.46)	**0.66 (0.50 - 0.86)**	0.50 (0.20 - 1.22)	C (vs. T/G)
TGFB1	rs1800470 [34]	A	UMNR	UMNR	0.95 (0.71 - 1.28)	1.17 (0.53 - 2.58)	0.94 (0.60 - 1.47)	0.92 (0.41 - 2.08)	C (vs. T/G)
TGFB2	rs4846480, Severe COPD[10]	C, AA	UMNR	UMNR	**1.26 (1.16 - 1.37)**				A (vs. T)
TGFB2	rs10429950 [25]	UMNR	UMNR	UMNR	**1.11 (1.07 - 1.14)**				T (vs. C)
THSD4	rs10851839 [42]	C	UMNR	UMNR	**1.10 (1.07 - 1.13)**				T (vs. A)
TIRAP	rs8177374 [24]	UMNR	UMNR	UMNR	**1.21 (1.12-1.30)**				T (vs. C)
THSD4	rs1441358 [25]	UMNR	UMNR	UMNR	**1.13 (1.10 - 1.16)**				G (vs. T)
TNF	rs80267959 [13]	C, A	UMNR	UMNR	**1.39 (1.01 - 1.90)**		1.39 (0.97 - 1.99)		A (vs. G)
TNF	rs80267959 [13]	C	UMNR	UMNR	1.15 (0.76 - 1.73)		1.17 (0.68 - 2.00)		A (vs. G)
TNF	rs80267959 [13]	A	UMNR	UMNR	**1.58 (1.04 - 2.42)**		1.59 (0.99 - 2.54)		A (vs. G)
TNS1	rs2571445 [42]	C	UMNR	UMNR	**1.07 (1.05 - 1.10)**				A (vs. G)
TRIP11	rs7155279 [42]	C	UMNR	UMNR	**1.05 (1.03 - 1.08)**				G (vs. T)

VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	**1.64 (1.09 - 2.48)**	1F-1F (vs. 1S-1F+1S-1S+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR	**1.73 (1.07 - 2.81)**	1F-1F (vs. 1S-1F+1S-1S+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR	1.44 (0.57 - 3.66)	1F-1F (vs. 1S-1F+1S-1S+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	0.74 (0.46 - 1.19)	2-2 (vs. 1F-1F+1S-1F+1S-1S+2-1F+2-1S)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR	0.71 (0.34 - 1.48)	2-2 (vs. 1F-1F+1S-1F+1S-1S+2-1F+2-1S)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR	0.83 (0.56 - 1.24)	2-2 (vs. 1F-1F+1S-1F+1S-1S+2-1F+2-1S)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	**0.78 (0.65 - 0.94)**	1F-1S (vs. 1F-1F+1S-1S+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR	**0.70 (0.55 - 0.89)**	1F-1S (vs. 1F-1F+1S-1S+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR	0.93 (0.69 - 1.24)	1F-1S (vs. 1F-1F+1S-1S+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	1.03 (0.84 - 1.25)	1S-1S (vs. 1F-1F+1S-1F+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR	0.87 (0.60 - 1.26)	1S-1S (vs. 1F-1F+1S-1F+2-1F+2-1S+2-2)

VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR	1.10 (0.87 - 1.39)	1S-1S (vs. 1F-1F+1S-1F+2-1F+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	0.83 (0.69 - 1.01)	2-1S (vs. 1F-1F+1S-1F+1S-1S+2-1F+2-2)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR	0.83 (0.63 - 1.10)	2-1S (vs. 1F-1F+1S-1F+1S-1S+2-1F+2-2)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR	0.82 (0.64 - 1.04)	2-1S (vs. 1F-1F+1S-1F+1S-1S+2-1F+2-2)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	1.24 (0.88 - 1.76)	1F-2 (vs. 1F-1F+1S-1F+1S-1S+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR	1.19 (0.79 - 1.80)	1F-2 (vs. 1F-1F+1S-1F+1S-1S+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR	1.34 (0.63 - 2.84)	1F-2 (vs. 1F-1F+1S-1F+1S-1S+2-1S+2-2)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	**1.27 (1.04 - 1.54)**	1F (vs. 1S+2)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR	**1.35 (1.04 - 1.75)**	1F (vs. 1S+2)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR	1.13 (0.83 - 1.54)	1F (vs. 1S+2)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR	0.90 (0.72 - 1.11)	2 (vs. 1F+1S)

VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR		0.88 (0.63 - 1.24)	2 (vs. 1F+1S)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR		0.90 (0.69 - 1.19)	2 (vs. 1F+1S)
VDBP	rs4588, rs7041 [45]	C, A	UMNR	UMNR		**0.86 (0.77 - 0.96)**	1S (vs. 1F+2)
VDBP	rs4588, rs7041 [45]	A	UMNR	UMNR		**0.76 (0.64 - 0.89)**	1S (vs. 1F+2)
VDBP	rs4588, rs7041 [45]	C	UMNR	UMNR		0.97 (0.83 - 1.13)	1S (vs. 1F+2)
VPS53	rs11247558 [20]	C, AA	S	M	0.96 (0.85 - 1.07)		C (vs. G)
VPS53	rs11247558 [20]	C, AA	S	F	**1.34 (1.22 - 1.46)**		C (vs. G)
VPS53	rs11656538 [20]	C, AA	S	M	1.03 (0.91 - 1.16)		A (vs. C)
VPS53	rs11656538 [20]	C, AA	S	F	**1.42 (1.28 - 1.55)**		A (vs. C)
VPS53	rs34001232 [20]	C, AA	S	M	1.02 (0.90 - 1.14)		A (vs. T)
VPS53	rs34001232 [20]	C, AA	S	F	**1.41 (1.28 - 1.54)**		A (vs. T)
VPS53	rs34469205 [20]	C, AA	S	M	0.99 (0.86 - 1.11)		A (vs. T)
VPS53	rs34469205 [20]	C, AA	S	F	**0.71 (0.58 - 0.84)**		A (vs. T)
VPS53	rs34729304 [20]	C, AA	S	M	1.02 (0.90 - 1.14)		C (vs. G)

VPS53	rs34729304 [20]	C, AA	S	F	**1.41 (1.28 - 1.53)**	C (vs. G)
VPS53	rs35716682 [20]	C, AA	S	M	0.93 (0.82 - 1.03)	A (vs. G/T)
VPS53	rs35716682 [20]	C, AA	S	F	**1.34 (1.22 - 1.45)**	A (vs. G/T)
VPS53	rs4968100 [20]	C, AA	S	M	0.99 (0.87 - 1.10)	A (vs. G)
VPS53	rs4968100 [20]	C, AA	S	F	**0.72 (0.59 - 0.84)**	A (vs. G)
VPS53	rs4968102 [20]	C, AA	S	M	1.00 (0.89 - 1.12)	T (vs. C)
VPS53	rs4968102 [20]	C, AA	S	F	**1.38 (1.25 - 1.51)**	T (vs. C)

Results of most comprehensive genetic factors with weak significant associations (OR of >0.5 and <2.0)

Legend: *Symbols*. *: if unspecified, it is an 'overall' estimate, including a mixed population of ethnicities, genders and/or smoking status. †: This heterozygous model comparison is WW vs. WM, rather than WM vs. MM. ‡: This estimate was inversed to fit the genetic model and comparison in relation to other estimates.

Abbreviations. Genders: F = Female, M = Male, UMNR = Unclear/mixed/not reported; Smoking status: Non-S = Non-smokers, S = Smokers, UMNR = Unclear/mixed/not reported; Ethnicities: AA = African American, A = Asian, Af = African, Ar = Arabian, C = Caucasian, E = European, E-A = East Asian, I = Indian, S-A = South Asian, UMNR = Unclear/mixed/not reported. Estimates in bold face indicate significant associations.

Genetic models. M = Mutant; W = Wildtype. Allelic model: M vs. W; Recessive model = MM vs. MW+WW; Dominant model = MM+MW vs. WW; Heterozygous model = MW vs. WW; Homozygous model = MM vs. WW; Overdominant model = MW vs. MM+WW.

ORs displayed bold are significant ORs

Supplementary Table 5

Gene	Variant / rs number	Ethni-city	Smo-king*	Gender*	Allelic	Recessive	Dominant	Hetero-zygous	Homo-zygous	Over-dominant	Other	Risk allele
ABLIM3	rs3839234 [42]	C	UMNR	UMNR	0.98 (0.95 - 1.02)							Del (vs. G)
ADAM33	rs2280089 (T+1)[60]	A	UMNR	UMNR	0.88 (0.72 - 1.07)	1.06 (0.62 - 1.81)	0.82 (0.57 - 1.19)		1.00 (0.58 - 1.71)			A (vs. G)
ADAM33	rs2787094 (V4)[60]	E, A	UMNR	UMNR	0.95 (0.76 - 1.19)	0.89 (0.63 - 1.25)	0.96 (0.72 - 1.27)		0.87 (0.58 - 1.31)			C (vs. G)
ADAM33	rs2787094 (V4)[60]	E	UMNR	UMNR	0.98 (0.77 - 1.23)	1.06 (0.77 - 1.45)	0.96 (0.71 - 1.28)		1.02 (0.74 - 1.42)			C (vs. G)
ADAM33	rs2787094 (V4)[60]	A	UMNR	UMNR	0.91 (0.55 - 1.51)	0.74 (0.42 - 1.30)	0.95 (0.50 - 1.79)		0.74 (0.34 - 1.60)			C (vs. G)
ADAM33	rs528557 (S2)[60]	E, A	UMNR	UMNR	0.97 (0.74 - 1.27)	1.05 (0.75 - 1.46)	0.97 (0.71 - 1.33)		1.01 (0.66 - 1.52)			G (vs. C/A)
ADAM33	rs528557 (S2)[60]	E	UMNR	UMNR	0.88 (0.62 - 1.24)	0.92 (0.55 - 1.54)	0.82 (0.55 - 1.22)		0.88 (0.43 - 1.80)			G (vs. C/A)
ADAM33	rs528557 (S2)[60]	A	UMNR	UMNR	1.05 (0.68 - 1.62)	1.19 (0.73 - 1.94)	1.08 (0.68 - 1.72)		1.13 (0.67 - 1.92)			G (vs. C/A)
ADRB2	rs1042714 [37]	C, A, Af	UMNR	UMNR				0.94 (0.69 - 1.24)	1.00 (0.80 - 1.25)			G (vs. C/T)
ADRB2	rs1800888 [37]	C, A, Af	UMNR	UMNR				1.17 (0.96 - 1.44)	2.57 (0.54 - 12.36)			T (vs. C)

AHNAK	rs2509961 [42]	C	UMNR	UMNR	1.01 (0.97 - 1.05)	T (vs. C)
BMP6	rs6924424 [42]	C	UMNR	UMNR	1.00 (0.97 - 1.04)	G (vs. T)
C1GALT1	rs10246303 [42]	C	UMNR	UMNR	1.02 (0.98 - 1.06)	T (vs. A)
CASC20/ BMP2	rs6140050 [42]	C	UMNR	UMNR	1.01 (0.97 - 1.05)	C (vs. A)
CCDC91	rs2348418 [42]	C	UMNR	UMNR	1.03 (1.00 - 1.05)	C (vs. T)
CDC7/TG FBR3	rs1192404 [42]	C	UMNR	UMNR	1.02 (0.97 - 1.08)	G (vs. A)
CHRM3	rs6688537 [42]	C	UMNR	UMNR	1.03 (0.99 - 1.07)	A (vs. C)
CISD3	rs11658500 [42]	C	UMNR	UMNR	1.04 (0.99 - 1.10)	A (vs. G)
DNLZ	rs10870202 [42]	C	UMNR	UMNR	1.02 (0.99 - 1.05)	C (vs. T)
EFCAB5	rs59835752 [42]	C	UMNR	UMNR	1.03 (0.97 - 1.09)	A (vs. Del)
EFEMP1	rs1430193 [42]	C	UMNR	UMNR	0.99 (0.96 - 1.01)	T (vs. A)

EPHX1	rs2234922 [2]	C	UMNR	UM NR	1.01 (0.96 - 1.06)					G (vs. A/T)
EPHX1	rs2234922 [2]	E-A	UMNR	UM NR	0.92 (0.79 - 1.08)					G (vs. A/T)
FGD6	rs113745635[42]	C	UMNR	UM NR	1.03 (0.99 - 1.08)					T (vs. C)
GSTP1	rs1695 [54]	C, A	UMNR	UM NR	1.07 (0.82 - 1.40)	1.51 (0.93 - 2.44)	1.00 (0.76 - 1.33)	0.92 (0.73 - 1.17)	1.50 (0.86 - 2.61)	G (vs. A)
GSTP1	rs1695 [54]	A	UMNR	UM NR	1.04 (0.74 - 1.46)	1.66 (0.91 - 3.03)	0.95 (0.66 - 1.35)	0.86 (0.64 - 1.15)	1.59 (0.81 - 3.09)	G (vs. A)
GSTP1	rs1695 [54]	C	UMNR	UM NR	1.13 (0.71 - 1.81)	1.39 (0.62 - 3.12)	1.11 (0.66 - 1.85)	1.04 (0.70 - 1.55)	1.48 (0.56 - 3.91)	G (vs. A)
GSTT1	Null/Wt [40]	C, A	UMNR	UM NR		1.00 (0.82 - 1.22)				Null (vs. Wt)
GSTT1	Null/Wt [40]	C	UMNR	UM NR		1.20 (0.63 - 2.29)				Null (vs. Wt)
GSTT1	Null/Wt [40]	A	UMNR	UM NR		0.93 (0.73 - 1.19)				Null (vs. Wt)
GSTT1	Null/Wt [52]	C, A	S	UM NR	1.05 (0.87 - 1.26)					Null (vs. Wt)
GSTT1	Null/Wt [52]	A	S	UM NR	1.07 (0.88 - 1.31)					Null (vs. Wt)

GSTT1	Null/Wt [52]	C, A	S	M	1.48 (0.90 - 2.43)					Null (vs. Wt)
GSTT1	Null/Wt [52]	C, A	S	F	1.42 (0.44 - 4.58)					Null (vs. Wt)
GSTT1	Null/Wt [23]	C, A	S	F					1.64 (0.73 – 3.69)	Null (vs. Wt)
HLA-DQB1	rs114544105 [42]	C	UMNR	UMNR	1.06 (0.99 - 1.13)					A (vs. G)
HSD17B12	rs4237643 [42]	C	UMNR	UMNR	1.02 (0.99 - 1.05)					T (vs. G)
IL1B	rs1143634 [51]	E-A, S-A, Ar	UMNR	UMNR		0.97 (0.32 - 2.88)	1.16 (0.78 - 1.73)	1.02 (0.33 - 3.12)		A (vs. G)
IL1B	rs1143634 [51]	E-A	UMNR	UMNR		0.33 (0.01 - 8.19)	1.16 (0.75 - 1.78)	0.35 (0.01 - 8.54)		A (vs. G)
IL4	rs2070874 [4]	UMNR	UMNR	UMNR	1.05 (0.79 - 1.38)					C (vs. T)
KANSL1	rs35524223 [42]	C	UMNR	UMNR	1.01 (0.94 - 1.08)					A (vs. T)
KCNJ2	rs6501431 [42]	C	UMNR	UMNR	0.98 (0.95 - 1.01)					C (vs. T)
LINC01467/LINC00911	rs1698268 [42]	C	UMNR	UMNR	1.01 (0.96 - 1.06)					T (vs. A)

LOC102723639	rs35506 [42]	C	UMNR	UMNR	0.99 (0.95 - 1.03)				T (vs. A)
LOC105369591	rs567508 [42]	C	UMNR	UMNR	1.04 (0.99 - 1.10)				G (vs. A)
LOC107984427	rs145729347 [42]	C	UMNR	UMNR	0.99 (0.92 - 1.06)				C (vs. G/A)
LST1	rs28986170 [42]	C	UMNR	UMNR	1.02 (0.93 - 1.11)				Del (vs. AA)
LTA	rs909253 [4]	UMNR	UMNR	UMNR	1.00 (0.78 - 1.28)				A (vs. G/T)
LTBP4	rs113473882 [42]	C	UMNR	UMNR	1.04 (0.93 - 1.17)				T (vs. C)
MECOM	rs1344555 [42]	C	UMNR	UMNR	1.01 (0.98 - 1.04)				T (vs. C)
MGA	rs72724130 [42]	C	UMNR	UMNR	1.01 (0.92 - 1.10)				T (vs. A)
MICAL3	rs11704827 [42]	C	UMNR	UMNR	1.03 (0.99 - 1.08)				A (vs. T)
MMP1	rs1799750 [59]	C, A	UMNR	UMNR	0.99 (0.89 - 1.10)	0.99 (0.81 - 1.21)	1.03 (0.86 - 1.22)	0.93 (0.77 - 1.12)	1C (vs. 2C)
MMP1	rs1799750 [59]	A	UMNR	UMNR	0.97 (0.67 - 1.41)	0.68 (0.39 - 1.20)	1.20 (0.75 - 1.92)	0.76 (0.40 - 1.44)	1C (vs. 2C)

MMP1	rs1799750 [59]	C	UMNR	UMNR	0.99 (0.92 - 1.07)	1.12 (0.96 - 1.31)	0.93 (0.82 - 1.06)		0.99 (0.85 - 1.16)	1C (vs. 2C)
MMP1	rs1799750 [59]	C, A	Smoking index matched between cases and controls	UMNR	1.00 (0.69 - 1.45)	0.80 (0.41 - 1.54)	1.17 (0.72 - 1.91)		0.85 (0.43 - 1.68)	1G (vs. 2G)
MMP3	rs3025058 [27]	C, A	UMNR	UMNR	0.95 (0.77 - 1.17)	0.74 (0.53 - 1.06)	1.14 (0.83 - 1.55)	1.41 (0.97 - 2.06)	0.85 (0.55 - 1.31)	6A (vs. 5A)
MMP3	rs35068180 [59]	C, A	UMNR	UMNR	0.88 (0.61- 1.27)	0.91 (0.58 - 1.41)	0.92 (0.58 - 1.46)		1.18 (0.76 - 1.82)	5A (vs. 6A/6C)
MMP12	rs2276109 [59]	C, A	UMNR	UMNR	0.98 (0.80 - 1.20)	0.96 (0.76 - 1.21)	1.14 (0.59 - 2.20)		1.17 (0.61 - 2.25)	T (vs. C)
MSRB3	rs1494502 [42]	C	UMNR	UMNR	1.03 (0.99 - 1.07)					A (vs. G)
MYPN	rs7095607 [42]	C	UMNR	UMNR	1.02 (0.98 - 1.06)					A (vs. G)
QSOX2	rs10858246 [42]	C	UMNR	UMNR	1.00 (0.98 - 1.03)					C (vs. G)
SERPINA3	rs4934 [4]	UMNR	UMNR	UMNR	0.80 (0.62 - 1.03)					A (vs. G/C)
SERPINE2	rs3795879 [41]	C, A	UMNR	UMNR	1.23 (0.97 - 1.32)	1.19 (0.85 - 1.66)	1.18 (0.85 - 1.62)	1.19 (0.81 - 1.76)[†]	1.23 (0.89 - 1.70)	C (vs. T/A)
SERPINE2	rs3795879 [41]	A	UMNR	UMNR	1.10 (0.89 - 1.36)	1.08 (0.86 - 1.37)	1.43 (0.70 - 2.91)	1.01 (0.83 - 1.34)[†]	1.45 (0.71 - 2.96)	C (vs. T/A)

SERPINE2	rs3795879 [41]	C	UMNR	UMNR	1.15 (0.92 - 1.45)	1.32 (0.64 - 2.73)	1.12 (0.78 - 1.60)	1.36 (0.58 - 3.24)[†]	1.17 (0.82 - 1.69)	C (vs. T/A)
SOD2	rs4880 [40]	C, A	UMNR	UMNR			1.07 (0.82 - 1.40)			T (vs. C)
SOD3	rs1799895 [40]	C, A	UMNR	UMNR			0.63 (0.25 - 1.60)			G (vs. C)
SFTPB	Combination of rs1130866 (G>A/C)and rs2077079 (G>T)[35]	C, A	UMNR	UMNR	1.13 (0.73 - 1.74)	1.18 (0.48 - 2.92)	2.02 (0.92 - 4.42)	1.10 (0.46 - 2.66)	2.38 (0.91 - 6.20)	W (vs. M)
SFTPD	Combination of rs1051246 (A>G), rs2245121 (G>A), rs2255601 (G>A), rs3088308 (T>A), rs6413520 (A>G), rs721917 (A>G)and rs911887 (T>C)[35]	C, A	UMNR	UMNR	1.09 (0.85 - 1.40)	1.07 (0.79 - 1.44)	1.07 (0.71 - 1.61)	1.06 (0.82 - 1.37)	1.06 (0.64 - 1.76)	W (vs. M)
SH3GL3	rs66650179 [42]	C	UMNR	UMNR	1.04 (0.97 - 1.12)					Del (vs. A)
SPAG17/ TBX15	rs200154334 [42]	C	UMNR	UMNR	1.00 (0.95 - 1.05)					CAT (vs. C
SUCLG2	rs1490265 [42]	C	UMNR	UMNR	1.01 (0.97 - 1.05)					C (vs. A)
SVIL/JCA D	rs3847402 [42]	C	UMNR	UMNR	1.01 (0.97 - 1.06)					A (vs. G)
TARS/LO C340113	rs91731 [42]	C	UMNR	UMNR	1.03 (0.96 - 1.09)					A (vs. C)
TGFB1	rs1800469 [34]	C, A	UMNR	UMNR	0.89 (0.77 - 1.02)	0.89 (0.72 - 1.11)	0.88 (0.66 - 1.18)		0.87 (0.66 - 1.14)	A (vs. G)

TGFB1	rs1800469 [34]	C	UMNR	UMNR	0.84 (0.68 - 1.05)					A (vs. G)
TGFB1	rs1800469 [34]	A	UMNR	UMNR	0.89 (0.77 - 1.02)	0.92 (0.73 - 1.17)	0.93 (0.64 - 1.35)	0.93 (0.69 - 1.26)		A (vs. G)
TGFB1	rs2241712 [34]	A	UMNR	UMNR	1.03 (0.89 - 1.20)	1.03 (0.52 - 2.07)	1.15 (0.84 - 1.58)	1.08 (0.45 - 2.58)		T (vs. C)
TGFB1	rs6957 [34]	C, A	UMNR	UMNR	1.14 (0.95 - 1.36)	1.45 (0.89 - 2.36)	1.30 (0.74 - 2.30)	1.52 (0.91 - 2.56)		T (vs. C)
TGFB1	rs6957 [34]	C	UMNR	UMNR	1.19 (0.92 - 1.54)					T (vs. C)
TGFB1	rs6957 [34]	A	UMNR	UMNR	1.02 (0.78 - 1.33)					T (vs. C)
TGFB1	rs2241718 [34]	C, A	UMNR	UMNR	0.95 (0.79 - 1.14)					G (vs. A)
TGFB1	rs2241718 [34]	A	UMNR	UMNR	0.91 (0.71 - 1.16)					G (vs. A)
TGFB2/MIR548F3	rs993925 [42]	C	UMNR	UMNR	1.02 (0.99 - 1.04)					C (vs. T)
TGFBR3	rs12140637 [42]	C	UMNR	UMNR	1.02 (0.98 - 1.06)					T (vs. C)
THSD4	rs12591467 [42]	C	UMNR	UMNR	1.03 (0.99 - 1.07)					C (vs. T)

TRAF3IP1 /ASB1	rs61332075 [42]	C	UMNR	UMNR	1.02 (0.96 - 1.08)		G (vs. C)
TSEN54/ CASKIN2	rs7218675 [42]	C	UMNR	UMNR	1.00 (0.96 - 1.04)		A (vs. C)
TIMP2	rs2277698 [4]	UMNR	UMNR	UMNR	0.59 (0.23 - 1.48)		A (vs. G)
TNF	rs1800610 [40]	C, A	UMNR	UMNR		1.10 (0.92 - 1.33)	A (vs. G)
TNF	rs361525 [12]	C, A	UMNR	UMNR	0.97 (0.69 - 1.37)	0.85 (0.59 - 1.21)	A (vs. G)
WWOX	rs1079572 [42]	C	UMNR	UMNR	1.00 (0.98 - 1.03)		A (vs. G)
ZGPAT	rs72448466 [42]	C	UMNR	UMNR	1.04 (1.00 - 1.08)		Del (vs. GT)
ZKSCAN1	rs72615157 [42]	C	UMNR	UMNR	1.02 (0.97 - 1.07)		G (vs. A)

Results of most comprehensive genetic factors with no significant associations.
Legend: *Symbols*. *: if unspecified, it is an 'overall' estimate, including a mixed population of ethnicities, genders and/or smoking status. †: This heterozygous model comparison is WW vs. WM, rather than WM vs. MM. ‡: This estimate was inversed to fit the genetic model and comparison in relation to other estimates. $: rs1130866 (C>T)and rs2077079 (C>A), rs1051246 (A>G), rs2245121 (G>A), rs2255601 (G>A), rs3088308 (T>A), rs6413520 (A>G), rs721917 (T>C)and rs911887 (T>C).

Abbreviations. Genders: F = Female, M = Male, UMNR = Unclear/mixed/not reported; Smoking status: Non-S = Non-smokers, S = Smokers, UMNR = Unclear/mixed/not reported; Ethnicities: AA = African American, A = Asian, Af = African, Ar = Arabian, C = Caucasian,

E = European, E-A = East Asian, I = Indian, S-A = South Asian, UMNR = Unclear/mixed/not reported. Estimates in bold face indicate significant associations.

Genetic models. M = Mutant; W = Wildtype. Allelic model: M vs. W; Recessive model = MM vs. MW+WW; Dominant model = MM+MW vs. WW; Heterozygous model = MW vs. WW; Homozygous model = MM vs. WW; Overdominant model = MW vs. MM+WW.

Supplementary Table 6

Risk factor	Comparison	OR	AR%*
Active smoking MC: Forey (2011) [64] Additionally in MMA: Wang (2015)[74] Jayes (2016)[66], and Kamal [67]	Ever vs. never smoking (any product)	*Overall*: **2.61 (2.03 - 3.19)(MMA; I^2=80%; n=3)[a]**	62
		Cohort: **2.82 (1.16 - 4.49)(MMA I^2=90.4%)[a]**	65
		Case-control: **2.49 (1.44 - 3.54)(MMA; I^2=45%; n=2)[a]**	60
		Cross-sectional: **2.54 (2.32 - 2.80)**	61
		Males: **2.77 (1.13 - 4.41)(MMA; I^2=91%; n=2)[a]**	64
		Females: **2.71 (2.22 - 3.20)(MMA; I^2=0%; n=2)[a]**	61
		European: **2.79 (2.46 - 3.16)**	64
		North American: **3.48 (2.88 - 4.20)**	71
		Asian: **2.80 (2.30 - 3.30)(MMA; I^2=0%; n=2)[a]**	64
	Current vs. never smoking (any product)	*Overall*: **3.51 (3.16 - 3.86)(MMA; I^2=0%; n=3)[b]**	72
		Cohort: **4.35 (3.27 - 5.43)(MMA; I^2=0%; n=2)[c]**	77
		Case-control: **4.69 (2.83 - 7.77)**	79
		Cross-sectional: **2.99 (2.67 - 3.31)(MMA; I^2=0%; n=2)[c]**	67
		Males: **4.03 (3.15 - 4.90)(MMA; I^2=0%; n=2)[d]**	75
		Females: **3.28 (2.35 - 4.58)**	70
		European: **3.69 (2.96 - 4.41)(MMA; I^2=27.1%; n=3)[b]**	73

	Non-European: **3.23 (2.44 - 4.02)(MMA; I^2=0%; n=5)**[b]	69
	American: **3.36 (1.06 - 5.66)(MMA; I^2=94%; n=2)**[c]	70
	Asian: **2.97 (2.38 - 3.56)(MMA I^2=0%; n=2)**[c]	66
Former vs. never smoking (any product)	*Overall*: **2.44 (2.15 - 2.73)(MMA; I^2=23%; n=2)**[c]	59
	Cohort: **2.89 (2.16 - 3.62)(MMA; I^2=0%; n=2)**[c]	65
	Case-control: **3.45 (2.26 - 5.28)**	71
	Cross-sectional: **2.26 (1.63 - 2.89)(MMA; I^2=80%; n=2)**[c]	56
	Males: **2.87 (2.35 - 3.50)**	65
	Females: **2.02 (1.53 - 1.68)**	50
	European: **2.20 (1.60 - 2.81)(MMA; I^2=57%; n=2)**[c]	55
	American: **2.64 (1.99 - 3.29)(MMA; I^2=42%; n=2)**[c†]	62
	Asian: **3.36 (2.28 - 4.44)(MMA I^2=86%; n=2)**[c]	70
About 5 cigs/day vs. never smoking	**2.89 (2.41 - 3.45)**	65
About 20 cigs/day vs. never smoking	**6.21 (4.72 - 8.17)**	84
About 45 cigs/day vs. never smoking	**9.50 (7.38 - 12.22)**	89
Highest vs. lowest amount of cigs/day	**2.32 (1.90 - 2.83)**	57
About age 26 years to start smoking vs. never smoking	**1.91 (1.25 - 1.91)**	48
About age 18 years to start smoking vs. never smoking	**2.11 (1.08 - 4.11)**	53
About age 14 years to start smoking vs. never smoking	**3.12 (2.07 - 4.70)**	68
Earliest vs. latest age at start smoking	**1.49 (1.26 - 1.76)**	33

	About 5 pack-years vs. never smoking	**1.25 (1.09 - 1.44)**	20
	About 20 pack-years vs. never smoking	**2.53 (1.87 - 3.43)**	60
	About 45 pack-years vs. never smoking	**3.69 (2.79 - 4.86)**	73
	Highest vs. lowest amount of pack-years	**2.80 (2.37 - 3.30)**	64
	Longest vs. shortest total duration of smoking	1.12 (0.63 - 1.98)	-‡
	About 3 years quit vs. never smoking	4.08 (0.80 - 20.77)	-‡
	About 7 years quit vs. never smoking	**4.94 (1.21 - 20.07)**	80
	About 12 years quit vs. never smoking	**2.12 (1.06 - 4.26)**	53
	Shortest vs. longest duration of quitting	**2.21 (1.24 - 3.94)**	55
	About 3 years quit vs. current smoking	0.77 (0.51 - 1.15)	-‡
	About 7 years quit vs. current smoking	1.03 (0.62 - 1.70)	-‡
	About 12 years quit vs. current smoking	**0.52 (0.37 - 0.71)**	-48
Dietary pattern MC: Zheng (2016)[76]	Highest vs. lowest category of intake of the healthy/prudent dietary pattern	**0.55 (0.46 - 0.66)**	-45
	Highest vs. lowest category of intake of the unhealthy/western-style dietary pattern	**2.12 (1.64 - 2.74)**	53
Passive smoking MC: Fischer (2015)[63] Additionally in MMA: Yang (2017)[73]	Second hand smoke exposed vs. not exposed	*Overall*: 1.32 (0.66 – 1.99)(MMA I^2=89%; n=2)[d]	-‡
		Females: **2.17 (1.48 - 3.18)**	54
Waterpipe tobacco smoking MC: Waziry (2017)[75]	Waterpipe tobacco smoking vs. no waterpipe tobacco smoking	**3.18 (1.25 - 8.08)**	69

Results of all most comprehensive lifestyle factors

Legend: *References belonging to MMA estimates:* a: this MMA is based on estimates from Forey (2011),[64] Wang (2015),[74] and Yang

(2017);[73] b: this MMA is based on estimates from Forey (2011),[64] Jayes (2016),[66] and Kamal (2015);[67] c: this MMA is based on estimates from Forey (2011)[64] and Kamal (2015);[67] d: this MMA is based on estimates from Fischer (2015)[63] and Yang (2017)[73]

Abbreviations: AR% = Attributable risk percent; JEM = Job Exposure Matrix; MC = Most comprehensive article selected; MMA = Meta-meta-analysis; OR = Odds Ratio

Symbols: * The AR% can be interpreted as the percentage of disease incidence among the exposed that are the result of the exposure, and therefore could be prevented if the exposure were eliminated. In case of protective factors (RR<1.00), the AR% can be interpreted as the percentage of cases that could be avoided if the entire population were exposed to this protective factor. † This estimate refers to the entire American continent ‡ No AR% could be calculated because the association was not significant.

ORs displayed bold are significant ORs.

Supplementary Table 7

Risk factor	Comparison	OR	AR%*
Living circumstances MC: Yang (2017) [73] for living circumstances	Living in a city vs. living in a town	1.23 (0.67 - 1.88)	-[†]
	Living around a polluted area	**1.63 (1.20 – 2.21)**	39
	Poor ventilation	**3.99 (1.24 – 12.82)**	75
Solid fuel smoke MC: Kurmi (2010) [68]	Any solid fuel smoke exposure vs. cooking on gas, oil or electricity	**2.80 (1.85 - 4.23)**	64
MC: Sana (2018)[72] MMA: Sana (2018)[72] and Yang[73]	Biomass smoke exposure vs. cooking on gas, oil or electricity	1.52 (0.59 – 2.45) (MMA; I^2=61.1%; n=2)[a]	-[†]
MC: Sana (2018) [72]		*Women*: 1.20 (0.99 – 1.40)	-[†]
MC: Kurmi (2010) [68]	Wood smoke exposure vs. cooking on gas, oil or electricity	*Women, non-smokers:* **1.80 (1.48 - 2.20)** **4.29 (1.35 - 13.70)**	44 77
Frequent cooking MC: Yang (2017) [73]	Cooking frequently	1.53 (0.79 – 2.93)	-[†]
Occupational exposure to vapours, gases, dust or fumes MC: Sadhra (2017)[71] Other used in MMA: Yang (2017) [73]	Occupational vs. no exposure to vapours, gases, dust or fumes	*Overall*: **1.22 (1.18 - 1.27)**	18
		Cohort: **1.11 (1.08 - 1.14)**	10
		Case-control: **1.75 (1.51 - 2.01)**	43
		Cross-sectional: **1.21 (1.13 - 1.29)**	17
		Male: **1.32 (1.21 - 1.45)**	24
		Female: **1.78 (1.42 - 2.23)**	44

		Determination of exposure status:	48
		Self-reported: **1.91 (1.72 - 2.13)**	
		Determination of exposure status:	10
		JEM-based: **1.10 (1.06 - 1.24)**	
	Low occupational vs. no exposure to vapours, gases, dust or fumes	0.77 (0.29 - 2.05)	-†
	Medium occupational vs. no exposure to vapours, gases, dust or fumes	1.07 (0.75-1.54)	-†
	High occupational vs. no exposure to vapours, gases, dust or fumes	**1.36 (1.14 - 1.63)**	26
	Occupational vs. no exposure to vapours	**1.24 (1.08 - 1.42)**	19
	Occupational vs. no exposure to gases	**1.10 (1.04 - 1.17)**	9
	Occupational vs. no exposure to dusts	**1.38 (1.29 - 1.47)(MMA; I^2=1%; n=2)[b]**	28
	Occupational vs. no exposure to biological dust	**1.33 (1.17 - 1.51)**	25
	Occupational vs. no exposure to mineral dust	**1.07 (1.05 - 1.09)**	7
	Occupational vs. no exposure to fumes	**1.16 (1.09 - 1.23)**	14
	Occupational vs. no exposure to fibres	1.76 (0.89 - 3.47)	-†
Physical labour work MC: Yang (2017) [73]	Physical labour work vs. no physical labour work	1.29 (0.81 – 2.07)	-†

Results of all most comprehensive environmental factors.
Legend. *References belonging to MMA estimates:* a: this MMA is based on estimates from Sana (2018)[72] and Yang;[73] b: this MMA is based on estimates from Yang (2017)[73] and Sadhra (2017).[71]

Abbreviations: AR% = Attributable risk percent; JEM = Job Exposure Matrix; MC = Most comprehensive article selected; MMA = Meta-meta-analysis; OR = Odds Ratio

Symbols: * The AR% can be interpreted as the percentage of disease incidence among the exposed that are the result of the exposure, and therefore could be prevented if the exposure were eliminated. In case of protective factors (RR<1.00), the AR% can be interpreted

as the percentage of cases that could be avoided if the entire population were exposed to this protective factor. † No AR% could be calculated because the association was not significant

ORs displayed bold are significant ORs.

Supplementary References

1. Aierken H, Wang J, Wushouer Q, et al. Polymorphisms of the ADAM33 gene and chronic obstructive pulmonary disease risk: A meta-analysis. *Clinical Respiratory Journal* 2014;8(1):108-15.
2. An L, Xia H, Zhou P, et al. Exploration of association between EPHX1 and chronic obstructive pulmonary disease on the basis of combined data mining. *Genetics and Molecular Research* 2016;15 (2) (no pagination)(15028639)
3. Brogger J, Steen VM, Eiken HG, et al. Genetic association between COPD and polymorphisms in TNF, ADRB2 and EPHX1. *European Respiratory Journal* 2006;27(4):682-88.
4. Castaldi PJ, Cho MH, Cohn M, et al. The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. *Human molecular genetics* 2010;19(3):526-34. doi: 10.1093/hmg/ddp519 [published Online First: 2009/11/26]
5. Chen H, Zhang L, He Z, et al. Vitamin D binding protein gene polymorphisms and chronic obstructive pulmonary disease: A meta-analysis. *Journal of Thoracic Disease* 2015;7(8):1423-40.
6. Chen L, Shen Y, Liu L, et al. Interleukin-13 -1112 C/T promoter polymorphism confers risk for COPD: a meta-analysis. *PloS one* 2013;8(7):e68222. doi: 10.1371/journal.pone.0068222 [published Online First: 2013/07/23]
7. Chen L, Wang T, Liu L, et al. Matrix Metalloproteinase-9 -1562C/T Promoter Polymorphism Confers Risk for COPD: A Meta-Analysis. *PloS one* 2013;8 (3) (no pagination)(e60523)
8. Cho MH, Boutaoui N, Klanderman BJ, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nature genetics* 2010;42(3):200.
9. Cho MH, Castaldi PJ, Wan ES, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Human molecular genetics* 2012;21(4):947-57.
10. Cho MH, McDonald ML, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The Lancet Respiratory medicine* 2014;2(3):214-25. doi: 10.1016/s2213-2600(14)70002-5 [published Online First: 2014/03/14]
11. Cui K, Ge X, Ma H. Four SNPs in the CHRNA3/5 alpha-neuronal nicotinic acetylcholine receptor subunit locus are associated with COPD risk based on meta-analyses. *PloS one* 2014;9(7):e102324. doi: 10.1371/journal.pone.0102324 [published Online First: 2014/07/23]
12. Cui K, Ge XY, Ma HL. Association of -238G/A and -863C/A polymorphisms in the TNF-alpha gene with chronic obstructive pulmonary disease based on a meta-analysis. *Genetics and Molecular Research* 2013;12(4):4981-89.
13. Cui K, Ge XY, Ma HL. Association of the TNF-alpha +489 G/A polymorphism with chronic obstructive pulmonary disease risk in Asians: Meta-analysis. *Genetics and Molecular Research* 2015;14(2):5210-20.
14. Dahl M, Hersh C, Ly N, et al. The protease inhibitor PI* S allele and COPD: a meta-analysis. *European Respiratory Journal* 2005;26(1):67-76.
15. Ding Z, Wang K, Li J, et al. Association between Glutathione S-transferase Gene M1 and T1 Polymorphisms and Chronic Obstructive Pulmonary Disease Risk: A Meta-analysis. *Clinical genetics* 2018
16. Du Y, Xue Y, Xiao W. Association of IREB2 Gene rs2568494 Polymorphism with Risk of Chronic Obstructive Pulmonary Disease: A Meta-Analysis. *Medical science monitor :*

international medical journal of experimental and clinical research 2016;22:177-82. [published Online First: 2016/01/19]
17. Duan L, Liang R, Wang Z, et al. Lack of association between the G+2044A polymorphism of interleukin-13 gene and chronic obstructive pulmonary disease: a meta-analysis. *Molecular biology reports* 2014;41(9):6297-303. doi: 10.1007/s11033-014-3512-z [published Online First: 2014/07/06]
18. Van Durme YMTA, Eijgelsheim M, Joos GF, et al. Hedgehog-interacting protein is a COPD susceptibility gene: The Rotterdam study. *European Respiratory Journal* 2010;36(1):89-95.
19. Gong Y, Fan L, Wan H, et al. Lack of association between the TGF-beta(1) gene and development of COPD in Asians: a case-control study and meta-analysis. *Lung* 2011;189(3):213-23. doi: 10.1007/s00408-011-9294-3 [published Online First: 2011/05/11]
20. Hardin M, Cho MH, Sharma SS, et al. Sex-based Genetic Association Study Identifies CELSR1 as a Possible COPD Risk Locus Among Women. *American journal of respiratory cell and molecular biology* 2016 doi: 10.1165/rcmb.2016-0172OC [published Online First: 2016/11/18]
21. Hersh C, Dahl M, Ly N, et al. Chronic obstructive pulmonary disease in α1-antitrypsin PI MZ heterozygotes: a meta-analysis. *Thorax* 2004;59(10):843-49.
22. Hu G, Shi Z, Hu J, et al. Association between polymorphisms of microsomal epoxide hydrolase and COPD: results from meta-analyses. *Respirology (Carlton, Vic)* 2008;13(6):837-50. doi: 10.1111/j.1440-1843.2008.01356.x [published Online First: 2008/09/25]
23. Hu G, Yao W, Zhou Y, et al. Meta- and pooled analyses of the effect of glutathione S-transferase M1 and T1 deficiency on chronic obstructive pulmonary disease. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2008;12(12):1474-81. [published Online First: 2008/11/20]
24. Hobbs BD, Parker MM, Chen H, et al. Exome array analysis identifies a common variant in IL27 associated with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2016;194(1):48-57.
25. Hobbs BD, de Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nature Genetics* 2017;49(3):426-32.
26. Ji G-Y, Tang X-Y, Wu S-Q, et al. Association between the IL-6 rs1800796 (-572C/G) polymorphism and chronic obstructive pulmonary disease risk: a meta-analysis. *Int J Clin Exp Med* 2017;10(7):10375-81.
27. Jiang S, Yang ZH, Chen YY, et al. MMP-9 genetic polymorphism may confer susceptibility to COPD. *Genetics and molecular research : GMR* 2016;15(2) doi: 10.4238/gmr.15026272 [published Online First: 2016/05/14]
28. Kang SW, Kim SK, Chung JH, et al. Genetic Polymorphism of Angiotensin-Converting Enzyme and Chronic Obstructive Pulmonary Disease Risk: An Updated Meta-Analysis. *BioMed Research International* 2016;2016 (no pagination)(7636123)
29. Lee J, Nordestgaard BG, Dahl M. EPHX1 polymorphisms, COPD and asthma in 47,000 individuals and in meta-analysis. *The European respiratory journal* 2011;37(1):18-25. doi: 10.1183/09031936.00012110 [published Online First: 2010/06/03]

30. Li DD, Guo SJ, Jia LQ, et al. Association of a disintegrin and metalloproteinase 33 (ADAM33) gene polymorphisms with chronic obstructive pulmonary disease in the Chinese population: A meta-analysis. *Genetics and Molecular Research* 2014;13(3):6391-97.
31. Li H, Fu WP, Hong ZH. Microsomal epoxide hydrolase gene polymorphisms and risk of chronic obstructive pulmonary disease: A comprehensive meta-analysis. *Oncology letters* 2013;5(3):1022-30. doi: 10.3892/ol.2012.1099 [published Online First: 2013/02/22]
32. Li W, Lan F, Yan F, et al. Angiotensin-converting enzyme I/D polymorphism is associated with COPD risk in Asian population: evidence from a meta-analysis. *Copd* 2013;10(1):35-9. doi: 10.3109/15412555.2012.727047 [published Online First: 2013/01/01]
33. Li X, Wei N, Wu Z, et al. The D/I polymorphism in the angiotensin-converting enzyme gene and chronic obstructive pulmonary disease risk: a meta-analysis. *Copd* 2012;9(5):485-91. doi: 10.3109/15412555.2012.694921 [published Online First: 2012/08/11]
34. Liao N, Zhao H, Chen M-L, et al. Association between the TGF-β1 polymorphisms and chronic obstructive pulmonary disease: a meta-analysis. *Biosci Rep* 2017;37(4) doi: 10.1042/BSR20170747
35. Ma T, Liu X, Liu Z. Functional polymorphisms in surfactant protein genes and chronic obstructive pulmonary disease risk: a meta-analysis. *Genetic testing and molecular biomarkers* 2013;17(12):910-7. doi: 10.1089/gtmb.2013.0308 [published Online First: 2013/10/08]
36. Nedeljkovic I, Terzikhan N, Vonk JM, et al. A genome-wide linkage study for chronic obstructive pulmonary disease in a Dutch genetic isolate identifies novel rare candidate variants. *Frontiers in genetics* 2018;9:133.
37. Nielsen AO, Jensen CS, Arredouani MS, et al. Variants of the ADRB2 Gene in COPD: Systematic Review and Meta-Analyses of Disease Risk and Treatment Response. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2017:1-10.
38. Niu LM, Liang Y, Xu M, et al. Effect of polymorphisms in the beta2-adrenergic receptor on the susceptibility and pulmonary function of patients with chronic obstructive pulmonary disease: a meta analysis. *Chinese medical journal* 2012;125(12):2213-8. [published Online First: 2012/08/14]
39. Smolonska J, Wijmenga C, Postma DS, et al. Erratum: Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. *Am J Respir Crit Care Med* 2010;181(7):765. doi: 10.1164/ajrccm.181.7.765 [published Online First: 2010/04/01]
40. Smolonska J, Wijmenga C, Postma DS, et al. Meta-analyses on suspected chronic obstructive pulmonary disease genes: A summary of 20 years' research. *American Journal of Respiratory and Critical Care Medicine* 2009;180(7):618-31.
41. Tian DB, Cai SX, Pan GC, et al. Association between the rs3795879 G/A polymorphism of the SERPINE2 gene and chronic obstructive pulmonary disease: A meta-analysis. *Genetics and Molecular Research* 2015;14(3):7920-28.
42. Wain LV, Shrine N, Artigas MS, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nature genetics* 2017;49(3):416.
43. Wang CD, Chen N, Huang L, et al. Impact of CYP1A1 Polymorphisms on Susceptibility to Chronic Obstructive Pulmonary Disease: A Meta-Analysis. *BioMed Research International* 2015;2015 (no pagination)(942958)

44. Wang YS, Liu L, Xu XY, et al. Association of interleukin-1beta -511C/T promoter polymorphism with COPD risk: a meta-analysis. *Genetics and molecular research : GMR* 2015;14(2):4477-84. doi: 10.4238/2015.May.4.5 [published Online First: 2015/05/13]
45. Wang YL, Kong H, Xie WP, et al. Association of vitamin D-binding protein variants with chronic obstructive pulmonary disease: a meta-analysis. *Genetics and molecular research : GMR* 2015;14(3):10774-85. doi: 10.4238/2015.September.9.16 [published Online First: 2015/09/25]
46. Wang W, Li P, Chen Y, et al. Association between beta2-Adrenergic Receptor-16Arg/Gly Gene Polymorphism and Chronic Obstructive Pulmonary Disease Risk:Systematic Review and Meta-Analysis. *Iranian journal of public health* 2014;43(7):877-88. [published Online First: 2015/04/25]
47. Xiao M, Guo L, Wang T, et al. Interleukin-1B-31T/C promoter polymorphism and chronic obstructive pulmonary disease risk: a meta-analysis. *Archives of medical science : AMS* 2014;10(3):434-8. doi: 10.5114/aoms.2014.43737 [published Online First: 2014/08/07]
48. Xiao M, Wang T, Zhu T, et al. Dual role of vitamin D-binding protein 1F allele in chronic obstructive pulmonary disease susceptibility: a meta-analysis. *Genetics and molecular research : GMR* 2015;14(2):3534-40. doi: 10.4238/2015.April.17.1 [published Online First: 2015/05/13]
49. Xie X, Zhang Y, Ke R, et al. Vitamin D-binding protein gene polymorphisms and chronic obstructive pulmonary disease susceptibility: A meta-analysis. *Biomedical reports* 2015;3(2):183-88. doi: 10.3892/br.2014.392 [published Online First: 2015/03/24]
50. Xie XM, Ke R, Zhang YH, et al. Interleukin-6 gene -174G>C polymorphism and chronic obstructive pulmonary disease risk: a meta-analysis. *Genetics and molecular research : GMR* 2015;14(3):8516-25. doi: 10.4238/2015.July.28.21 [published Online First: 2015/09/09]
51. Xie ZK, Huang QP, Huang J, et al. Association between the IL1B, IL1RN polymorphisms and COPD risk: a meta-analysis. *Scientific reports* 2014;4:6202.
52. Xue H, Su J, Sun K, et al. Glutathione S-transferase M1 and T1 gene polymorphism and COPD risk in smokers: an updated analysis. *Molecular biology reports* 2012;39(4):5033-42. doi: 10.1007/s11033-011-1300-6 [published Online First: 2011/12/14]
53. Yan F, Chen C, Jing J, et al. Association between polymorphism of glutathione S-transferase P1 and chronic obstructive pulmonary disease: a meta-analysis. *Respiratory medicine* 2010;104(4):473-80. doi: 10.1016/j.rmed.2010.01.009 [published Online First: 2010/02/02]
54. Yang L, Li X, Tong X, et al. Association between glutathione S-transferase P1 Ile (105) Val gene polymorphism and chronic obstructive pulmonary disease: A meta-analysis based on seventeen case-control studies. *Meta Gene* 2015;6:59-64.
55. Zhang J, Summah H, Zhu YG, et al. Nicotinic acetylcholine receptor variants associated with susceptibility to chronic obstructive pulmonary disease: a meta-analysis. *Respiratory research* 2011;12:158. doi: 10.1186/1465-9921-12-158 [published Online First: 2011/12/20]
56. Zhang L, Chang WW, Ding H, et al. Transforming growth factor-beta1 polymorphisms and chronic obstructive pulmonary disease: A meta-analysis. *International Journal of Tuberculosis and Lung Disease* 2011;15(10):1301-07.
57. Zhang L, Gu H, Gu Y, et al. Association between TNF-alpha -308 G/A polymorphism and COPD susceptibility: a meta-analysis update. *International journal of chronic obstructive*

pulmonary disease 2016;11:1367-79. doi: 10.2147/copd.s105394 [published Online First: 2016/07/07]
58. Zhang R, Li H, Zhao H, et al. Polymorphisms in a disintegrin and metalloprotease 33 gene and the risk of chronic obstructive pulmonary disease: A meta-analysis. *Respirology (Carlton, Vic)* 2014;19(3):312-20.
59. Zhou H, Wu Y, Jin Y, et al. Genetic polymorphism of matrix metalloproteinase family and chronic obstructive pulmonary disease susceptibility: a meta-analysis. *Scientific reports* 2013;3:2818.
60. Zhou DC, Zhou CF, Toloo S, et al. Association of a disintegrin and metalloprotease 33 (ADAM33) gene polymorphisms with the risk of COPD: an updated meta-analysis of 2,644 cases and 4,804 controls. *Molecular biology reports* 2015;42(2):409-22. doi: 10.1007/s11033-014-3782-5 [published Online First: 2014/10/05]
61. Zhou H, Ying X, Liu Y, et al. Genetic polymorphism of heme oxygenase 1 promoter in the occurrence and severity of chronic obstructive pulmonary disease: a meta-analysis. *Journal of cellular and molecular medicine* 2017;21(5):894-903.
62. Alif SM, Dharmage SC, Bowatte G, et al. Occupational exposure and risk of chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Expert Review of Respiratory Medicine* 2016;10(8):861-72. doi: 10.1080/17476348.2016.1190274
63. Fischer F, Kraemer A. Meta-analysis of the association between second-hand smoke exposure and ischaemic heart diseases, COPD and stroke. *Bmc Public Health* 2015;15 doi: 10.1186/s12889-015-2489-4
64. Forey BA, Thornton AJ, Lee PN. Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema. *Bmc Pulmonary Medicine* 2011;11 doi: 10.1186/1471-2466-11-36
65. Hu G, Zhou Y, Tian J, et al. Risk of COPD from exposure to biomass smoke: a metaanalysis. *Chest Journal* 2010;138(1):20-31.
66. Jayes L, Haslam PL, Gratziou CG, et al. SmokeHaz Systematic Reviews and Meta-analyses of the Effects of Smoking on Respiratory Health. *Chest* 2016;150(1):164-79. doi: 10.1016/j.chest.2016.03.060
67. Kamal R, Srivastava AK, Kesavachandran CN. Meta-analysis approach to study the prevalence of chronic obstructive pulmonary disease among current, former and non-smokers. 2015
68. Kurmi OP, Semple S, Simkhada P, et al. COPD and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis. *Thorax* 2010;65(3):221-28. doi: 10.1136/thx.2009.124644
69. Po JYT, FitzGerald JM, Carlsten C. Respiratory disease associated with solid biomass fuel exposure in rural women and children: systematic review and meta-analysis. *Thorax* 2011;66(3):232-39. doi: 10.1136/thx.2010.147884
70. Ryu JY, Sunwoo YE, Lee SY, et al. Chronic Obstructive Pulmonary Disease (COPD) and Vapors, Gases, Dusts, or Fumes (VGDF): A Meta-analysis. *Copd-Journal of Chronic Obstructive Pulmonary Disease* 2015;12(4):374-80. doi: 10.3109/15412555.2014.949000
71. Sadhra S, Kurmi OP, Sadhra SS, et al. Occupational COPD and job exposure matrices: a systematic review and meta-analysis. *International journal of chronic obstructive pulmonary disease* 2017;12:725-34. doi: 10.2147/copd.s125980

72. Sana A, Somda SM, Meda N, et al. Chronic obstructive pulmonary disease associated with biomass fuel use in women: a systematic review and meta-analysis. *BMJ Open Respiratory Research* 2018;5(1):e000246.
73. Yang Y, Mao J, Ye Z, et al. Risk factors of chronic obstructive pulmonary disease among adults in Chinese mainland: A systematic review and meta-analysis. *Respiratory Medicine* 2017;131:158-65.
74. Wang BY, Xiao D, Wang C. Smoking and chronic obstructive pulmonary disease in Chinese population: a meta-analysis. *Clinical Respiratory Journal* 2015;9(2):165-75. doi: 10.1111/crj.12118

Chapter 6

Network analysis of genome-wide association studies for chronic obstructive pulmonary disease in the context of biological pathways

Authors:

Sarah Mount*[1], Elisa Cirillo*[2], Kelly Stewart[3,4], Susan Coort[2], Chris T. Evelo[2], Anke Wesselius[3], Maurice P. Zeegers[3,4], Annemie Schols[1]

[1] Department of Respiratory Medicine, School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre Maastricht, Maastricht, The Netherlands

[2] Department of Bioinformatics - BiGCaT, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, the Netherlands

[3] Department of Complex Genetics and Epidemiology, School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre, Maastricht, The Netherlands

[4] CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre, Maastricht, The Netherlands

*Authors contributed equally/shared first authorship

Abstract

The aetiology of chronic obstructive pulmonary disease remains poorly understood. Meta-analyses and genome-wide association studies are common methods for studying genetic risk factors for disease. Although these studies are useful for the identification of candidate genes and variants, the interpretation of the results within a disease and pathway context remains challenging. In this study, we used network analysis to elucidate the biological context of the genetic variants Methods: In associated with chronic obstructive pulmonary disease and to perform analysis based on data visualization. We used data collected in a comprehensive review, including all known (181 statistically significant) chronic obstructive pulmonary disease related variants. Different visualizations of the same SNP-gene-pathway network were created in Cytoscape to investigate both the gene functionality and the potential variant effect in context. For the interpretation of the gene's roles and their connections in the network, we identified eleven functional classes obtained after the analysis on the specific gene function. Subsequently, we performed a variant effect predictor analysis to examine the influence of the significant single nucleotide polymorphisms on their respective genes. Significant genes were found in 315 biological pathways from the curated human collection of WikiPathways. We focused the analysis on the potentially influential seven missense single nucleotide polymorphisms present in different genes: *AK9, SERPINA1, IL27, CYP1A1, EPHX1, SLC22A11 and TESMIN*. Our results suggest that inflammatory and detoxification pathways may be the most relevant targets for future research on chronic obstructive pulmonary disease. This emphasises the relevance of gene-environment, gene-behaviour and gene-lifestyle interactions.

Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities.[1] [1]Unfortunately, COPD is the third leading cause of death worldwide[2] but on the other hand, it is largely preventable, as its primary risk factor is smoking.[3] Although, not all smokers will develop COPD,[4] and cases also occur in never smokers.[5] Given this marked variability in susceptibility, COPD risk may be attributable to genetic predispositions[6] and/or to environmental factors such as poor diet,[7 8] lack of exercise,[9] exposure to pollution, vapours, gases, infection[10] and/ or gene-environment interactions.[11] However, the causal mechanisms behind COPD remain poorly understood[12 13] and current treatments are limited to symptom control.[14] Determining the mechanism(s) behind the development of COPD, specifically which genetic factors are associated with it, could lead to better and more personalized interventions, prevention, and risk prediction, by allowing us to more specifically intervene with damaged gene-products.

As a result of genome-wide association studies (GWAS), significant advances have been made in the identification of risk loci for COPD. These studies have discovered many genetic variations, including single nucleotide polymorphisms (SNPs), which may be relevant to the development of COPD, but many of these SNPs are in non-coding regions of DNA,[15] or the gene function is not known. However, disease manifestations are rarely a result of single gene alteration, instead they reflect disturbances within complex intracellular networks.[16] Nonetheless, such large caches of information are difficult to interpret, particularly within the context of other systems, within the disease process itself and in the presence of unknown gene functions. Considering genes in context, as well as the gene variant effect aids in understanding the function of the included SNPs within biological contexts. Network analysis provides a way of deciphering the biological relationships between SNPs, genes and pathways, by providing a framework which allows for the integration, analysis and display of this complex data.[17 18] The creation of a SNP-gene-pathway network, does not only provide useful visualizations of the different types of data, but may also give new insights into the pathobiology of COPD.[19]

In this study we used data and knowledge-driven methods based on pathway and network analysis to examine the inter-relationships between the genes associated with COPD risk, identified through genetic variation associations with COPD. We made functional maps of the network to better understand the genes roles in the context of COPD as well as examined the types of variants and their biological consequences.

Methods

Dataset

The genetic variations analysed in this study have been taken from a comprehensive overview that was performed by Mount and Stewart. The authors performed a systematic search of Embase, Pubmed, and Web of Science to find and extract all genetic variants published in pooled and meta-analysis studies related to the genetic risk of COPD. A detailed description of study methods can be found on Prospero (CRD4201705) May 2018.

From this review we extracted the 181 significant genetic variants mapped to 99 genes. The list of variants includes 176 SNPs with an rs identifier and eight variations represented by: a combination of multiple SNPs, insertions and deletions or length polymorphisms [refer to Table E1). The latter two are included only in the SNP-gene-pathway network visualizations as nodes, but could not be further investigated due to the lack of a unique identifier. All genetic variants from the comprehensive overview have been included in our networks, regardless of any linkage disequilibrium.

Analysis workflow

The analysis workflow is presented in Figure 1, and shows the different steps and tools used. In the first step, the list of 181 variants and 99 genes were represented in a SNP-gene network using Cytoscape version 3.6,[20] an open source, modular, software platform for integrating, analysing and visualizing complex networks. In the second step, the genes were used to retrieve the biological pathways from the complete WikiPathways Human collection, version 20180710 (10 July 2018).[21] WikiPathways is an open and community-curated biological pathway database that allows for the analysis, visualization and integration of transcriptomic, proteomic, epigenomic, metabolomic and interaction related data.[21] Genes present in one or

more pathways were displayed in a Cytoscape gene-pathway network. In the third step, the SNP-gene and gene-pathway network were consolidated using the core "merge" function in Cytoscape. This yielded a SNP-gene-pathway network that was used as a basic reference for the biological interpretation of the connected elements. In the fourth step, additional investigations on the gene and SNPs descriptions were performed.

The retrieved information was included in the network visualizations and presented using table legends and/or colour coding. The SNP-gene-pathway networks are publicly available for consultation and further exploration at the NDEx (Network Data Exchange) website (http://www.ndexbio.org/),[22] and include the additional SNPs and gene information.

Results

Different visualization for the SNP - gene - pathway networks

Our analysis (Figure 1) produced four different visualizations of the same SNP-gene-pathway network. Each network visualization highlights different types of information and contains an attribute table related to the network node, for ease of interpretation. In Table 1, an overview of the main characteristics of the networks is reported. In each network title, the link to the NDEx visualization is provided, and the main features of the networks and nodes codes are reported.

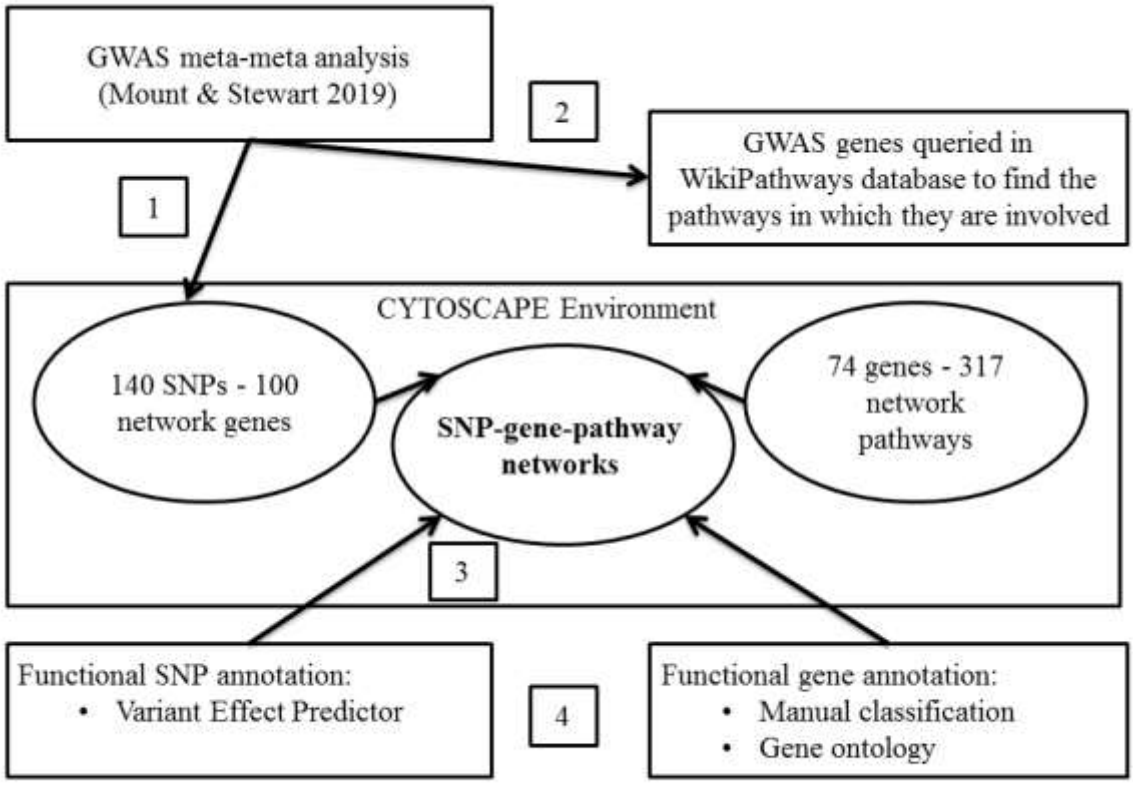


Figure 1: work flow of the analysis.

Different visualization for the SNP - gene - pathway networks

Our analysis (Figure 1) produced four different visualizations of the same SNP-gene-pathway network. Each network visualization highlights different types of information and contains an attribute table related to the network node, for ease of interpretation. In Table 1, an overview of the main characteristics of the networks is reported. In each network title, the link to the NDEx visualization is provided, and the main features of the networks and nodes codes are reported.

Table 1

Network title	Weblink	Features	Node colour
Gene - pathway network	http://doi.org/10.18119/N9SW2N	Result of the merged "SNP-gene" and "gene-pathway" networks. This is the basic set for the other two network visualizations.	SNP= green; gene= blue; pathway = red.
Functional-gene - map	http://doi.org/10.18119/N9JC76	Genes belonging to a specific functional class listed and colour coded.	SNP = grey triangle; gene= squared coloured node according to the; pathway = grey circle;
Variant - effect network	http://doi.org/10.18119/N9P301	SNPs with different type of gene impact are highlighted.	SNP low=green; SNP moderate = blue; SNP modifier = orange; gene=grey triangle; pathway = grey circle

Overview of the four SNP-gene-pathway networks

The "Gene - pathway network" represents the basic workflow output. All the networks account for 181 variants nodes, 99 genes nodes and 315 pathways nodes for a total number of 739 connections within the three elements. Of the original set of 99 genes, 74 genes are present in pathways from the curated WikiPathways. The basic version highlights the three elements: SNPs, genes and pathways in different colours. The other networks present more elaborate visualizations and are discussed in the following sections.

Gene function interpretation

The "Functional - gene - map", visualization presents "Functional classes" classification (Table E1) network. In this table, 13 non-overlapping functional classes are identified: Addiction (6), Cellular interaction (3), Cellular metabolism (17), Cellular structure (5), Detoxification (5),

Development (12), Homeostasis organismal (3), Inflammation (15), Lung function (2), Metabolism organismal (3), Regulation (11), Tissue remodelling (12), and Unknown (4), (refer to Figure E2 for coding legend). When examining the gene distribution in the network and pathway connections, it is interesting to observe that in some of the functional classes, genes are dispersed, while in others all are connected. Cellular metabolism (shown in dark green) is an example of this dispersion, 15 genes are not connected in the major central network and seven of the 15 do not present any pathway connections. Comparatively, all five genes related to Detoxification cluster in a specific area (displayed in pink). Similarly, all 15 genes involved in Inflammation are grouped in the central area of the major network (shown in purple), which are also intensely connected to other genes and pathways. This network visualization thus gives an indication of the biological process in which the 26 unlinked genes are involved. However, we relied on the additional information we collected, shown in the attribute table of the network (Table E2), for the biological interpretation.

Variant effect interpretation

The "Variant - effect network" visualization makes use of VEP Ensembl tool and gives an indication of the impact of the variants on the gene sequence (refer to Table E4). The resulting network contained 149 modifier (or non-coding variants) SNPs, 19 moderate (or missense) SNPs and two low (or synonymous) impact SNPs. Polyphen[23] and SIFT[24] prediction scores, presented in the network table as attributes, were also consulted to elucidate the deleteriousness of the 19 missense variants. The Polyphen resulted in five probably damaging SNPs (rs141159367, rs1051740, rs10499052, rs146043252, rs28929474,) and two possibly damaging (rs1048943, rs181206) SNPs. However, SIFT indicated that only three (rs1051740, rs10499052, rs146043252) of the seven SNPs identified by Polyphen are deleterious. Due to this discrepancy, we considered the more extensive list in further discussion and analysis. In Figure 2 a SNP-gene-pathway subnetwork is presented, highlighting only the connections related to the seven genes (*e.g. AK9, SERPINA1, IL27, CYP1A1, EPHX1, SLC22A11 and TESMIN*) carrying the deleterious missense SNPs.

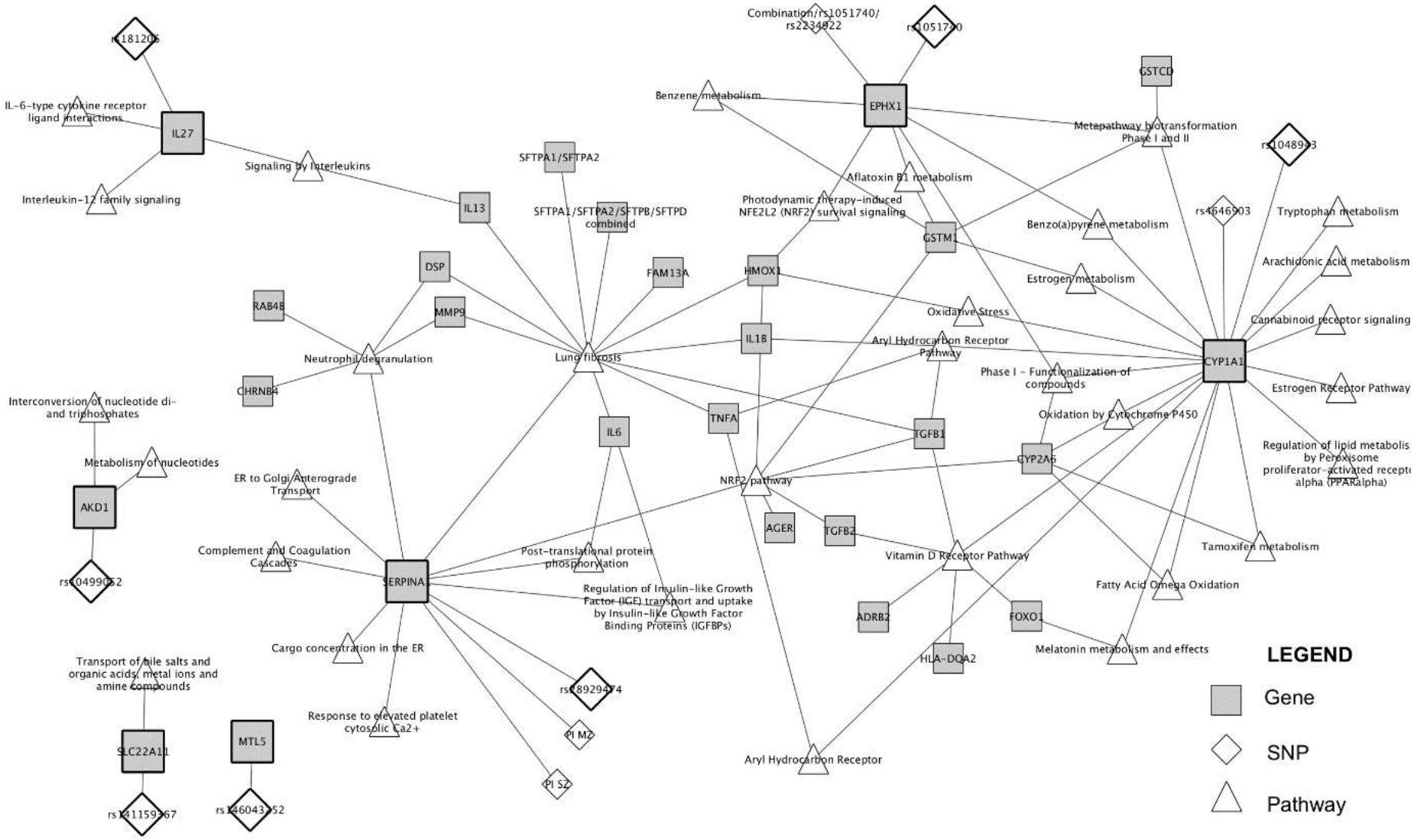


Figure 2: Subnetwork of the SNP-gene-pathway network highlighting the connection of the seven genes (*e.g. AK9, SERPINA1, IL27, CYP1A1, EPHX1, SLC22A11 and TESMIN*) carrying deleterious missense SNPs.

Discussion

We applied network analysis to genes associated with COPD to improve the interpretability of genetic variants. We provide an integrated view of the variants, the genes, their pathways and interactions. Most of the genes within a functional class grouped together, indicating that they share common pathways; while other genes tended to be more dispersed, perhaps dictated by the pleiotropic role of the genes.

Detoxification, inflammation, tissue remodelling and lung function genes are likely to be the most interesting to discuss with respect to COPD, as they may provide direct targets for risk modification. In contrast, the genes related to cellular metabolism, organismal metabolism, organismal homeostasis, cell structure, development and addiction pathways, although relevant, may be involved in more indirect causal pathways. The distinctions have been made in the functional classes based on the (probable) function of the genes. For example, genes classified in cellular metabolism class play roles in RNA or protein coding, genes with the function of maintaining energy are part of spliceosome metabolism class, and genes involved in nutrient and waste product balance are clustered in the organismal metabolism class. These classifications are subjective, but help to guide on the important functions of the gene/protein product in the context of COPD.

The VEP analysis presented here showed that the vast majority of the SNPs associated with COPD are modifiers, with a small number of missense mutations. Modifier SNPs may affect regulatory mechanisms, such as gene splicing, transcription factor binding, or messenger RNA degradation [25]. However, the influence of modifiers is likely limited, particularly if a single modifier SNP is considered, although they may be cumulatively relevant.[26]

Nineteen of the 181 variants were missense SNPs mutations. Of these, seven SNPs: rs10499052, rs28929474, rs181206, rs1048943, rs1051740, rs141159367 and rs146043252 showed deleterious alterations in the associated proteins for SLC22A11, AK9, SERPINA1, IL27, CYP1A1, EPHX1 and TESMIN respectively. Figure 2 displays the interactions of those seven genes with the missense SNPs and pathways. Interestingly, all but two of the deleterious

alterations, located in AK9 and TESMIN, are in genes which are either directly or indirectly, involved in inflammatory pathways. The AK9 gene mutation is involved in cellular metabolic processes and is primarily expressed in extra-pulmonary tissues,[27] while TESMIN is involved in heavy metal ion binding and sequestering.

The IL27 and SERPINA1 encode proteins directly involved in inflammation. The cytokine IL27 has both anti-inflammatory and inflammatory functions[28] while SERPINA1 encodes alpha-antitrypsin which acts to prevent the uncontrolled proteolytic attack of the lungs.[29] The leucine to proline substitution caused by rs181206 (IL27) was predicted to be possibly damaging by Polyphen, indicating a strong change in protein structure. Proline is known to have an exceptional conformational rigidity, often causing structural changes. This is in line with the results from a previous meta-analysis and clinical findings. Hobbs et. al. reported an increased OR of 1.18 in presence of proline (originally reported OR inversed to show the effect of the minor/risk allele).[30] Indeed, higher levels of IL27 have also been observed in the sputum and plasma of COPD patients, compared to healthy controls.[31] Moreover, levels have been observed to be further elevated (in serum) during exacerbations, and sputum levels are also negatively correlated with FEV1.[31 32] Lastly, cigarette smoke upregulates the naïve CD4+ T cell expression of IL27.[33] Together, these findings highlight the relevance of IL27 in COPD pathogenesis. Similar to IL27 in the inflammatory pathway is SERPINA1. This is perhaps the most well-known gene associated with COPD.[34] However, variants in this gene account for less than two per cent of all COPD cases.[35] Interestingly, the deleterious variant (PiMZ) associated with this gene was only observed in crude estimates and disappeared after adjusting for smoking behaviour.[36] Although this is puzzling, it could be attributable to other environmental or behavioural factors. To add to this puzzle, the findings of cohort studies remain inconsistent with respect to lung function decline in smoking and non-smoking individuals with the PiMZ variant.[37-40]

Detoxification as well as the oxidative stress may determine the susceptibility to COPD, but the mechanism of this risk remains poorly understood.[41] Interestingly we observed that two deleterious SNPs rs1048943 and rs1051740 respectively, associated with CYP1A1 and EPHX1,

may play a key role in this type of susceptibility. Both of these genes are involved in detoxification pathways, and under some circumstances, may be directly linked, as in Benzo(a)pyrene metabolism (WikiPathways identifier: WP696). Polycyclic aromatic hydrocarbons (PAH), found in cigarette smoke, air pollution,[42] coal tar, and grilled meats,[43] are catalysed by cytochrome P450 enzymes.[44] The risk associated with CYP1A1 (rs1048943) gene variant (462 Ile/Val) was 3.23 (1.50n-6.93),[45] and has been associated with markers of increased oxidative stress.[41] This stress is due to the CYP1A1 variant's ability to increase enzyme activity, and further activate tobacco carcinogens.[46] This is particularly relevant with respect to PAH metabolism, as CYP1A1 acts sequentially with EPHX1. The EPHX1 gene encodes a phase II biotransformation enzyme in epoxide metabolism.[47] The EPHX1 rs1051740 variant is known to reduce enzyme activity and is sometimes referred to as the slow allele.[48] The slow variant of this gene has been associated with an increased risk of COPD in Caucasians[49] while the fast variant (rs2234922) has been observed to be protective in Asians.[50] However, the contribution of the EPHX1 variants to COPD risk remains controversial.[51] If these variants were present in the same individual, i.e. the fast CYP1A1 variant, causing a build-up of carcinogens, and the slow EPHX1 variant, a build-up of dangerous xenobiotics and/or xenobiotic metabolites could arise thereby increasing the risk of COPD.

Finally, the SLC22A11 (rs141159367) gene is involved in the excretion of xenobiotics, endogenous organic anions, and urate.[52] The mutation results in an alanine to valine substitution. However, both are non-reactive and hydrophobic, thus this substitution may have an overall neutral affect.[53]

To our knowledge, this is the first time this type of network analysis has been performed on COPD GWAS meta-analytic data. The most comparable study was performed in 70 patients with emphysema and bronchiolitis.[54] However, our results deviated strongly from their findings, with the only similarities being commonly expressed risk altering genes being HMOX1, IL6, and AGER. These deviations may stem from differences in tissue collection,[55] our inclusion of only statistically relevant variants, or from differences in patient populations.

Strengths and Limitations

The methods demonstrated here have a number of strengths and limitations. We have used data collected at the highest level of evidence, a so-called meta-meta-analysis. This data is then presented within the pathways of disease, through the use of network analysis. This allows us to study the disease more precisely by putting candidate genes and gene variants into their pathway and interaction contexts, instead of only considering individual genetic determinants.[56] However, all genetic variants from the comprehensive overview have been included in our networks, regardless of any linkage disequilibrium. As a result, the discussed pathways, genes and variants may be affected by known and unknown linkage disequilibrium (LD). Lastly, by making our network publicly available in NDEX, we have made it available for further research and made it easier to reproduce our analysis.

Conclusion

In this study, we graphically presented SNPs, genes and pathways involved in the risk of COPD in extensive networks. Furthermore, we compartmentalized the genes into functional classes in order to simplify the visual representation of pathways and aid the understand-ability of this complex information. We also discuss the impact of seven deleterious missense gene-variants while explaining how to interpret the networks. The visual representation of the complex data, highlights relevant pathways, as well as targets for laboratory and human intervention trials.

Declarations

Conflict of Interest

None.

Author contributions

Conception and design: SM, EC, KS, SC, CT, AW, MZ, AS; Analysis and interpretation: SM, EC, KS; Drafting the manuscript for important intellectual content: SM, EC "; Revision: SM, EC, KS, SC, CT, AW, MZ, AS; Final approval: : SM, EC, KS, SC, CT, AW, MZ, AS.

Funding

This study was funded by the Maastricht University Interfaculty Program 'Eatwell'.

Acknowledgements

We would like to thank dr. Ramon Langen and dr. Niki Reynaert for helping us with the classification of the genes.

References

1. Agustí AG, Celli BR, Chen R, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary 2018 report: Global Intiative for Chronic Obstructive Lung Disease; 2018 [Available from: https://goldcopd.org accessed 6-7-2018 2018.
2. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet* 2012;380(9859):2095-128.
3. Raherison C, Girodet P. Epidemiology of COPD. *European Respiratory Review* 2009;18(114):213-21.
4. Bossé Y. Updates on the COPD gene list. *International journal of chronic obstructive pulmonary disease* 2012;7:607.
5. Lamprecht B, McBurnie MA, Vollmer WM, et al. COPD in never smokers: results from the population-based burden of obstructive lung disease study. *Chest* 2011;139(4):752-63.
6. Schamberger AC, Mise N, Meiners S, et al. Epigenetic mechanisms in COPD: implications for pathogenesis and drug discovery. *Expert opinion on drug discovery* 2014;9(6):609-28.
7. Varraso R, Shaheen SO. Could a healthy diet attenuate COPD risk in smokers?: BMJ Publishing Group Ltd, 2017.
8. Varraso R, Chiuve SE, Fung TT, et al. Alternate Healthy Eating Index 2010 and risk of chronic obstructive pulmonary disease among US women and men: prospective study. *bmj* 2015;350:h286.
9. Garcia-Aymerich J, Lange P, Benet M, et al. Regular physical activity modifies smoking-related lung function decline and reduces risk of chronic obstructive pulmonary disease: a population-based cohort study. *American journal of respiratory and critical care medicine* 2007;175(5):458-63.
10. Epidemiology of chronic obstructive pulmonary disease: prevalence, morbidity, mortality, and risk factors. Seminars in respiratory and critical care medicine; 2015. Thieme Medical Publishers.
11. de Jong K, Vonk JM, Timens W, et al. Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1 levels. *Journal of Allergy and Clinical Immunology* 2015;136(6):1664-72. e14.
12. Celli BR, Decramer M, Wedzicha JA, et al. An official American Thoracic Society/European Respiratory Society statement: research questions in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2015;191(7):e4-e27.
13. Barnes PJ, Burney PGJ, Silverman EK, et al. Chronic obstructive pulmonary disease. *Nature Reviews Disease Primers* 2015;1:15076. doi: 10.1038/nrdp.2015.76
14. Organization WH. Chronic obstructive pulmonary disease (COPD) Geneva, Switzerland: World Health Organization; 2018 [updated 15 November 2016. Available from: http://www.who.int/en/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) accessed July 6 2018.
15. Ortega VE. Picking the Right Fruit: Intersecting Chronic Obstructive Pulmonary Disease Genome-Wide Association Study Discoveries with Epigenetics: American Thoracic Society, 2018.
16. Abraham G, Inouye M. Genomic risk prediction of complex human disease and its clinical application. *Current opinion in genetics & development* 2015;33:10-16.

17. Diez D, Agustí A, Wheelock CE. Network analysis in the investigation of chronic respiratory diseases. From basics to application. *American journal of respiratory and critical care medicine* 2014;190(9):981-88.
18. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nature Reviews Genetics* 2010;11(12):843.
19. Faner R, Agustí A. Network analysis: a way forward for understanding COPD multimorbidity: Eur Respiratory Soc, 2015.
20. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003;13(11):2498-504.
21. Slenter DN, Kutmon M, Hanspers K, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research* 2017;46(D1):D661-D67.
22. Pillich RT, Chen J, Rynkov V, et al. NDEx: a community resource for sharing and publishing of biological networks. Protein Bioinformatics: Springer 2017:271-301.
23. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods* 2010;7(4):248.
24. Sim N-L, Kumar P, Hu J, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research* 2012;40(W1):W452-W57.
25. Nott A, Meislin SH, Moore MJ. A quantitative analysis of intron effects on mammalian gene expression. *Rna* 2003;9(5):607-17.
26. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Human molecular genetics* 2015;24(R1):R102-R10.
27. Panayiotou C, Solaroli N, Karlsson A. The many isoforms of human adenylate kinases. *The international journal of biochemistry & cell biology* 2014;49:75-83.
28. Wynick C, Petes C, Gee K. Interleukin-27 mediates inflammation during chronic disease. *Journal of Interferon & Cytokine Research* 2014;34(10):741-49.
29. Ekeowa UI, Marciniak SJ, Lomas DA. α1-antitrypsin deficiency and inflammation. *Expert review of clinical immunology* 2011;7(2):243-52.
30. Hobbs BD, Parker MM, Chen H, et al. Exome array analysis identifies a common variant in IL27 associated with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2016;194(1):48-57.
31. Cao J, Zhang L, Li D, et al. IL-27 is elevated in patients with COPD and patients with pulmonary TB and induces human bronchial epithelial cells to produce CXCL10. *Chest* 2012;141(1):121-30.
32. Angata T, Ishii T, Gao C, et al. Association of serum interleukin-27 with the exacerbation of chronic obstructive pulmonary disease. *Physiological reports* 2014;2(7)
33. Qiu S-L, Duan M-C, Liang Y, et al. Cigarette smoke induction of interleukin-27/WSX-1 regulates the differentiation of Th1 and Th17 cells in a smoking mouse model of emphysema. *Frontiers in immunology* 2016;7:553.
34. Židzik J, Slabá E, Joppa P, et al. Glutathione S-transferase and microsomal epoxide hydrolase gene polymorphisms and risk of chronic obstructive pulmonary disease in Slovak population. *Croatian medical journal* 2008;49(2):182-91.
35. Stoller JK, Aboussouan LS. α1-antitrypsin deficiency. *The Lancet* 2005;365(9478):2225-36.
36. Hersh C, Dahl M, Ly N, et al. Chronic obstructive pulmonary disease in α1-antitrypsin PI MZ heterozygotes: a meta-analysis. *Thorax* 2004;59(10):843-49.

37. Thun G-A, Ferrarotti I, Imboden M, et al. SERPINA1 PiZ and PiS heterozygotes and lung function decline in the SAPALDIA cohort. *PLoS One* 2012;7(8):e42728.
38. Dahl M, Tybjærg-Hansen A, Lange P, et al. Change in lung function and morbidity from chronic obstructive pulmonary disease in α1-antitrypsin MZ heterozygotes: a longitudinal study of the general population. *Annals of internal medicine* 2002;136(4):270-79.
39. Wadsworth ME, Vinall LE, Jones AL, et al. Alpha1-antitrypsin as a risk for infant and adult respiratory outcomes in a national birth cohort. *American journal of respiratory cell and molecular biology* 2004;31(5):559-64.
40. Silva GE, Sherrill DL, Guerra S, et al. A longitudinal study of α1-antitrypsin phenotypes and decline in FEV1 in a community population. *Chest* 2003;123(5):1435-40.
41. Vibhuti A, Arif E, Mishra A, et al. CYP1A1, CYP1A2 and CYBA gene polymorphisms associated with oxidative stress in COPD. *Clin Chim Acta* 2010;411(7-8):474-80. doi: 10.1016/j.cca.2009.12.018
42. Ewa B, Danuta M-Š. Polycyclic aromatic hydrocarbons and PAH-related DNA adducts. *Journal of applied genetics* 2017;58(3):321-30.
43. Information NCfB. Benzo(A)Pyrene Bethesda, USA: National Center for Biotechnology Information, U.S. National Library of Medicin; [updated 2018-07-21. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/2336 accessed July 26 2018.
44. Sibiriak S, Sergeeva S, Kurchatova N, et al. Effect of rifampicin, endotoxin, and their combination on the phagocyte activity, microsomal oxidation, and free radical processes in the liver. *Eksperimental'naia i klinicheskaia farmakologiia* 1997;60(6):34-36.
45. Wang C-D, Chen N, Huang L, et al. Impact of CYP1A1 polymorphisms on susceptibility to chronic obstructive pulmonary disease: a meta-analysis. *BioMed research international* 2015;2015
46. Autrup H. Genetic polymorphisms in human xenobiotica metabolizing enzymes as susceptibility factors in toxic response. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 2000;464(1):65-76.
47. Omiecinski CJ, Hassett C, Hosagrahara V. Epoxide hydrolase—polymorphism and role in toxicology. *Toxicology letters* 2000;112:365-70.
48. Penyige A, Poliska S, Csanky E, et al. Analyses of association between PPAR gamma and EPHX1 polymorphisms and susceptibility to COPD in a Hungarian cohort, a case-control study. *BMC medical genetics* 2010;11(1):152.
49. Smolonska J, Wijmenga C, Postma DS, et al. Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. *American journal of respiratory and critical care medicine* 2009;180(7):618-31.
50. Guoping H, Zhe S, Jinxing H, et al. Association between polymorphisms of microsomal epoxide hydrolase and COPD: Results from meta-analyses. *Respirology* 2008;13(6):837-50.
51. Lakhdar R, Denden S, Knani J, et al. Combined analysis of EPHX1, GSTP1, GSTM1 and GSTT1 gene polymorphisms in relation to chronic obstructive pulmonary disease risk and lung function impairment. *Disease markers* 2011;30(5):253-63.
52. Sakiyama M, Matsuo H, Shimizu S, et al. A common variant of organic anion transporter 4 (OAT4/SLC22A11) gene is associated with renal underexcretion type gout. *Drug metabolism and pharmacokinetics* 2014;29(2):208-10.

53. Betts MJ, Russell RB. Amino acid properties and consequences of substitutions. Bioinformatics for geneticists2003:289-316.
54. Faner R, Cruz T, Casserras T, et al. Network analysis of lung transcriptomics reveals a distinct B-cell signature in emphysema. *American journal of respiratory and critical care medicine* 2016;193(11):1242-53.
55. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2011;7(1):32-43.
56. Silverman EK, Loscalzo J. Network medicine approaches to the genetics of complex diseases. *Discovery medicine* 2012;14(75):143.

Supplementary methods

Functional annotation of identified genes

In order to classify the genes presented in the merged network into functional classes, we collected and reported Human Gene Name Consortium (HGNC) approved symbols and gene acronyms (Table E1). Next, the function of the gene was examined using the gene descriptions in GeneCards (https://www.genecards.org) (reported as a link), as well as secondary sources including PubMed. Using this information, we determined what the general function of the gene was assigned a "Synthetic function" term, such as neurotransmitter, homeostatic control, enzyme, a transmembrane protein, etc. Then, we used this information in order to make a "General classification" for the gene depending on the general function of the gene such as neurotransmission, metabolism or cellular interactions. Furthermore, we identified a "Functional classes", a term which summarizes the genes' functionalities within the disease context using the previous information, as well as further consulting PubMed and GeneCards, (Table E2). In the event that the gene function could not be determined from the above sources, a Gene Ontology (GO) predicted term was assigned using ARCHS4 (https://amp.pharm.mssm.edu/archs4/). This is a database which estimates gene functions using GO terms. However, no information was available on two genes (*CRAMP1L* and *EPB41L4A-AS1*), thus their functions and classification remain unknown. Although they may possess regulator roles since the latter is a non-coding antisense RNA, while the former is a cramped chromatin regulator homolog. The "Functional class" terms obtained, were used to colour the gene nodes in the SNP-gene-network.

As a fourth step, we performed a GO enrichment analysis using GO-Elite E(1) to ensure that we captured the most important functionalities of the genes. Lastly, experts in molecular

pathways and COPD pathology were consulted to verify the classifications of the genes and gene pathways.

Gene ontology analysis

The GO enrichment analysis was performed using GO-Elite. The 99 genes from the comprehensive review were annotated with GO terms using GO-Elite E(1) in which 203 terms were detected in the three top-level GO trees: molecular function, biological process and cellular component. The GO annotation of these genes was obtained running the GO-Elite analysis with default parameters: Z-score cut off for initial filtering above 1.96, three minimum number of genes changed genes connected to a GO term, permuted p-value cut off above 0.05, excluding terms with gene ID counts greater than 10000. The pruned results list was consulted and reported in Table E3.

Functional annotation of variants

In the set of the 181 variants, 140 were SNPs with an annotated rs identifier. The description of the 140 SNPs was retrieved using Variant Effect Predictor (VEP) analysis E(2). The tool provides different types of variant descriptions on genes, transcripts, and protein sequences as well as regulatory regions. In addition, it indicates the potential variant effect on the gene or sequence in which the SNP is located. The settings chosen for the analysis are shown in Figure E1, and the output with the information for the different SNPs is available in Table E4. The SNPs are classified into modifier, moderate and low impact SNPs, which were visualized in the SNP-gene-pathway network with different colour According to VEP, these three classifications refer to the type of consequence of a variant sequence on the gene. Modifier is the classification for variants located in non-coding regions both intra and inter-genic E(3),

moderate is a term used to indicate missense variants changing the protein sequence of the gene, and low impact variants are synonymous SNPs that do not have any consequence on the amino acid sequence of the protein. The VEP analysis also provides the Polyphen and SIFTS prediction scores that were consulted to elucidate the deleteriousness of the missense variants. The SIFT or the sorting intolerant from tolerant E(4) is an algorithm which uses sequence homology to predict whether an amino acid substitution affects protein function E(5). PolyPhen or Polymorphism phenotyping calculates the impact of amino acid alterations on the stability and function of proteins using structural and comparative evolutionary considerations E(6).

Supplementary figures

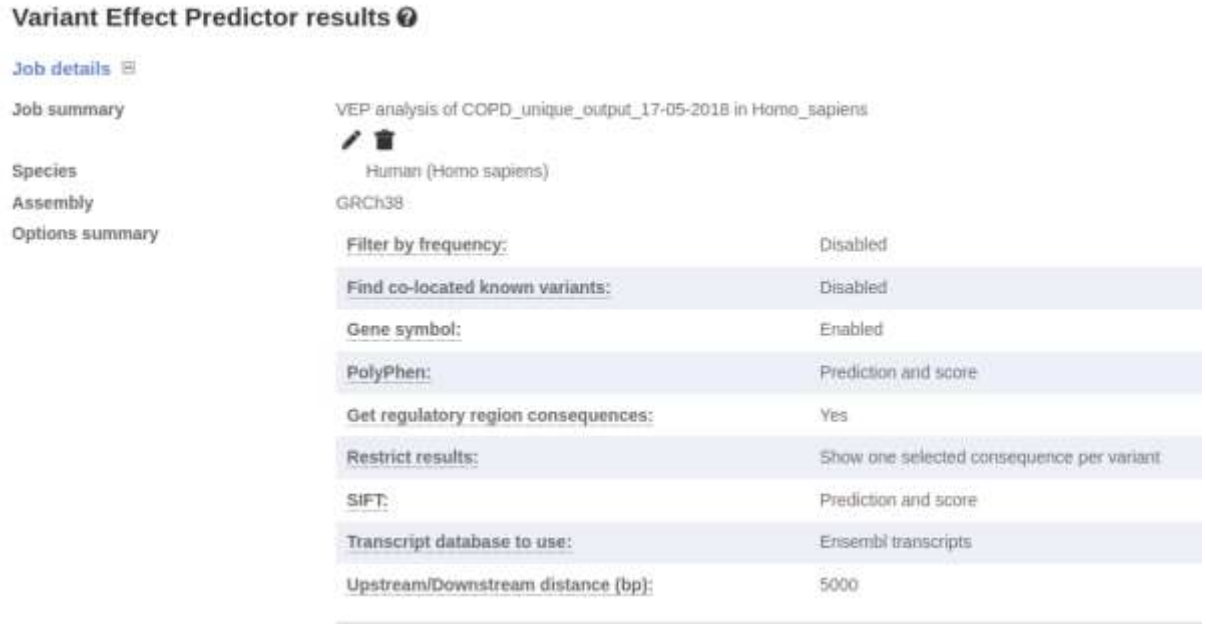


Figure E1 : VEP settings

Column	Functional Network area
Mapping Type	Discrete Mapping
Addiction	R:51 G:255 B:255 - #33FFFF
Cellular-interaction	R:51 G:255 B:51 - #33FF33
Cellular-metabolism	R:0 G:102 B:0 - #006600
Cellular-structure	R:204 G:0 B:0 - #CC0000
Detoxification	R:255 G:153 B:255 - #FF99FF
Development	R:153 G:153 B:153 - #999999
Homeostasis- organismal	R:255 G:0 B:153 - #FF0099
Inflammation	R:204 G:0 B:204 - #CC00CC
Lung function	R:0 G:0 B:255 - #0000FF
Metabolism-organismal	R:153 G:0 B:51 - #990033
Regulation	R:255 G:255 B:51 - #FFFF33
Tissue-remodelling	R:255 G:153 B:51 - #FF9933
Unknown	R:0 G:0 B:0 - #000000

Figure E2: Legend of the colour code corresponding to a specific “Functional classes” term.

Supplementary tables

Table E1

Gene name	Gene Ensembl Identifier	Variant / rs number	Outcome
ACE	ENSG00000159640	del/ins	COPD
ADAM19	ENSG00000135074	rs113897301	COPD
ADAM33	ENSG00000149451	rs2280090	COPD
ADAM33	ENSG00000149451	rs2280090	COPD
ADAM33	ENSG00000149451	rs2280091	COPD
ADAM33	ENSG00000149451	rs2280091	COPD
ADAM33	ENSG00000149451	rs3918396	COPD
ADAM33	ENSG00000149451	rs3918396	COPD
ADAM33	ENSG00000149451	rs511898	COPD
ADAM33	ENSG00000149451	rs597980	COPD
ADAM33	ENSG00000149451	rs612709	COPD
ADAM33	ENSG00000149451	rs612709	COPD
ADAM33	ENSG00000149451	rs612709	COPD
ADGRG6	ENSG00000112414	rs9399401	COPD
ADRB2	ENSG00000169252	rs1042713	COPD
AGER	ENSG00000204305	rs2070600	COPD
HYKK	ENSG00000188266	rs8042849	COPD
HYKK	ENSG00000188266	rs8042849	COPD
HYKK	ENSG00000188266	rs9788721	COPD
HYKK	ENSG00000188266	rs9788721	COPD
AKD1	ENSG00000155085	rs10499052	COPD
ARMC2	ENSG00000118690	rs2806356	COPD
CCDC101	ENSG00000176476	rs17707300	COPD
CELSR1	ENSG00000075275	rs56344079	COPD
CELSR1	ENSG00000075275	rs7286446	COPD
CELSR1	ENSG00000075275	rs7286446	COPD
CELSR1	ENSG00000075275	rs9615358	COPD
CELSR1	ENSG00000075275	rs9615973	COPD
CELSR1	ENSG00000075275	rs9615981	COPD
CELSR1	ENSG00000075275	rs9615981	COPD
CELSR1	ENSG00000075275	rs9615982	COPD
CELSR1	ENSG00000075275	rs9615982	COPD
CFDP1	ENSG00000153774	rs7186831	COPD
CHRNA3	ENSG00000080644	rs1051730	COPD
CHRNA3	ENSG00000080644	rs1051730	COPD

CHRNA3	ENSG00000080644	rs1051730	COPD
CHRNA3	ENSG00000080644	rs1051730	COPD
CHRNA3	ENSG00000080644	rs114205691	COPD
CHRNA3	ENSG00000080644	rs114205691	COPD
CHRNA3	ENSG00000080644	rs12914385	COPD
CHRNA3	ENSG00000080644	rs12914385	COPD
CHRNA3	ENSG00000080644	rs12914385	Moderate-to-severe COPD
CHRNA3	ENSG00000080644	rs12914385	Severe COPD
CHRNA3	ENSG00000080644	rs138544659	COPD
CHRNA3	ENSG00000080644	rs138544659	COPD
CHRNA3	ENSG00000080644	rs141518190	COPD
CHRNA3	ENSG00000080644	rs141518190	COPD
CHRNA3	ENSG00000080644	rs146009840	COPD
CHRNA3	ENSG00000080644	rs146009840	COPD
CHRNA3	ENSG00000080644	rs147144681	COPD
CHRNA3	ENSG00000080644	rs147144681	COPD
CHRNA3	ENSG00000080644	rs147499554	COPD
CHRNA3	ENSG00000080644	rs147499554	COPD
CHRNA3	ENSG00000080644	rs4887067	COPD
CHRNA3	ENSG00000080644	rs4887067	COPD
CHRNA3	ENSG00000080644	rs55676755	COPD
CHRNA3	ENSG00000080644	rs55676755	COPD
CHRNA3	ENSG00000080644	rs56077333	COPD
CHRNA3	ENSG00000080644	rs56077333	COPD
CHRNA3	ENSG00000080644	rs6495309	COPD
CHRNA3	ENSG00000080644	rs8034191	COPD
CHRNA3	ENSG00000080644	rs8192482	COPD
CHRNA3	ENSG00000080644	rs8192482	COPD
CHRNA5	ENSG00000169684	rs11633958	COPD
CHRNA5	ENSG00000169684	rs11633958	COPD
CHRNA5	ENSG00000169684	rs140330585	COPD
CHRNA5	ENSG00000169684	rs140330585	COPD
CHRNA5	ENSG00000169684	rs16969968	COPD
CHRNA5	ENSG00000169684	rs16969968	COPD
CHRNA5	ENSG00000169684	rs16969968	COPD
CHRNA5	ENSG00000169684	rs17486195	COPD
CHRNA5	ENSG00000169684	rs17486195	COPD
CHRNA5	ENSG00000169684	rs17486278	COPD
CHRNA5	ENSG00000169684	rs17486278	COPD

CHRNA5	ENSG00000169684	rs17486278	COPD
CHRNA5	ENSG00000169684	rs190065944	COPD
CHRNA5	ENSG00000169684	rs190065944	COPD
CHRNA5	ENSG00000169684	rs2036527	COPD
CHRNA5	ENSG00000169684	rs2036527	COPD
CHRNA5	ENSG00000169684	rs55853698	COPD
CHRNA5	ENSG00000169684	rs55853698	COPD
CHRNA5	ENSG00000169684	rs56390833	COPD
CHRNA5	ENSG00000169684	rs56390833	COPD
CHRNA5	ENSG00000169684	rs7172118	COPD
CHRNA5	ENSG00000169684	rs7172118	COPD
CHRNA5	ENSG00000169684	rs7180002	COPD
CHRNA5	ENSG00000169684	rs7180002	COPD
CHRNA5	ENSG00000169684	rs72740955	COPD
CHRNA5	ENSG00000169684	rs72740955	COPD
CHRNA5	ENSG00000169684	rs72740964	COPD
CHRNA5	ENSG00000169684	rs72740964	COPD
CHRNA5	ENSG00000169684	rs951266	COPD
CHRNA5	ENSG00000169684	rs951266	COPD
CHRNB4	ENSG00000117971	rs55988292	COPD
CHRNB4	ENSG00000117971	rs55988292	COPD
CHRNB4	ENSG00000117971	rs72743158	COPD
CHRNB4	ENSG00000117971	rs72743158	COPD
CHRNB4	ENSG00000117971	rs17487223	COPD
CHRNB4	ENSG00000117971	rs17487223	COPD
CRAMP1L	ENSG00000007545	rs61746451	COPD
CYP1A1	ENSG00000140465	rs1048943	COPD
CYP1A1	ENSG00000140465	rs4646903	COPD
CYP1A1	ENSG00000140465	rs4646903	COPD
CYP2A6	ENSG00000255974	rs12459249	COPD
DSP	ENSG00000096696	rs2076295	COPD
EEFSEC	ENSG00000132394	rs2811416	COPD
EEFSEC	ENSG00000132394	rs2811416	COPD
EEFSEC	ENSG00000132394	rs2955083	COPD
EPB41L4A-AS1	ENSG00000224032	rs66669542	COPD
EPHX1	ENSG00000143819	Combination of rs1051740 and rs2234922	COPD
EPHX1	ENSG00000143819	Combination of rs1051740 and rs2234922	COPD

EPHX1	ENSG00000143819	Combination of rs1051740 and rs2234922	COPD
EPHX1	ENSG00000143819	Combination of rs1051740 and rs2234922	COPD
EPHX1	ENSG00000143819	rs1051740	COPD
EPHX1	ENSG00000143819	rs1051740	COPD
FAM13A	ENSG00000138640	rs10021465	COPD
FAM13A	ENSG00000138640	rs10021465	COPD
FAM13A	ENSG00000138640	rs1812329	COPD
FAM13A	ENSG00000138640	rs1812329	COPD
FAM13A	ENSG00000138640	rs1964516	COPD
FAM13A	ENSG00000138640	rs2013701	COPD
FAM13A	ENSG00000138640	rs2013701	COPD
FAM13A	ENSG00000138640	rs2045517	COPD
FAM13A	ENSG00000138640	rs2045517	COPD
FAM13A	ENSG00000138640	rs28455964	COPD
FAM13A	ENSG00000138640	rs28455964	COPD
FAM13A	ENSG00000138640	rs2869966	COPD
FAM13A	ENSG00000138640	rs2869966	COPD
FAM13A	ENSG00000138640	rs2869967	COPD
FAM13A	ENSG00000138640	rs2869967	COPD
FAM13A	ENSG00000138640	rs2904259	COPD
FAM13A	ENSG00000138640	rs2904259	COPD
FAM13A	ENSG00000138640	rs3846287	COPD
FAM13A	ENSG00000138640	rs3846287	COPD
FAM13A	ENSG00000138640	rs3857043	COPD
FAM13A	ENSG00000138640	rs3857043	COPD
FAM13A	ENSG00000138640	rs4416442	COPD
FAM13A	ENSG00000138640	rs4416442	COPD
FAM13A	ENSG00000138640	rs4416442	Moderate-to-severe COPD
FAM13A	ENSG00000138640	rs4693980	COPD
FAM13A	ENSG00000138640	rs4693980	COPD
FAM13A	ENSG00000138640	rs6830970	COPD
FAM13A	ENSG00000138640	rs6830970	COPD
FAM13A	ENSG00000138640	rs6837671	COPD
FAM13A	ENSG00000138640	rs6837671	COPD
FAM13A	ENSG00000138640	rs6837671	COPD
FAM13A	ENSG00000138640	rs76273989	COPD
FAM13A	ENSG00000138640	rs7671167	COPD
FAM13A	ENSG00000138640	rs7671167	COPD

FAM13A	ENSG00000138640	rs7671167	COPD
FAM13A	ENSG00000138640	rs7671261	COPD
FAM13A	ENSG00000138640	rs7671261	COPD
FAM13A	ENSG00000138640	rs7674369	COPD
FAM13A	ENSG00000138640	rs7674369	COPD
FAM13A	ENSG00000138640	rs7682317	COPD
FAM13A	ENSG00000138640	rs7682317	COPD
FAM13A	ENSG00000138640	rs7682431	COPD
FAM13A	ENSG00000138640	rs7682431	COPD
FAM208B	ENSG00000108021	rs41290259	COPD
FAR2	ENSG00000064763	rs7294481	COPD
FAR2	ENSG00000064763	rs7294481	COPD
FOXO1	ENSG00000150907	rs75700692	COPD
FOXO1	ENSG00000150907	rs78372177	COPD
GEMIN4	ENSG00000179409	rs11652959	COPD
GSTCD	ENSG00000138780	rs11727735	COPD
GSTM1	ENSG00000134184	Null/Wt	COPD
GSTM1	ENSG00000134184	Null/Wt	COPD
GSTM1	ENSG00000134184	Null/Wt	COPD
GSTM1	ENSG00000134184	Null/Wt	COPD
GSTM1	ENSG00000134184	Null/Wt	COPD
GSTM1	ENSG00000134184	Null/Wt	COPD
GYPA	ENSG00000170180	rs13105210	COPD
GYPA	ENSG00000170180	rs13105210	COPD
GYPA	ENSG00000170180	rs4835177	COPD
GYPA	ENSG00000170180	rs4835177	COPD
HHIP	ENSG00000164161	rs13118928	COPD
HHIP	ENSG00000164161	rs13141641	COPD
HMOX1	ENSG00000100292	Length polymorphism	COPD
HMOX1	ENSG00000100292	Length polymorphism	COPD
HMOX1	ENSG00000100292	Length polymorphism	COPD
HMOX1	ENSG00000100292	Length polymorphism	COPD
HMOX1	ENSG00000100292	Length polymorphism	COPD
HTR4	ENSG00000164270	rs7733088	COPD
IL1B	ENSG00000125538	rs1143627	COPD
IL1B	ENSG00000125538	rs16944	COPD
IL6	ENSG00000136244	rs1800795	COPD
IL6	ENSG00000136244	rs1800796	COPD
IL13	ENSG00000169194	rs1800925	COPD

IL13	ENSG00000169194	rs1800925	COPD
IL13	ENSG00000169194	rs20541	COPD
IL13	ENSG00000169194	rs20541	COPD
IL27	ENSG00000197272	rs181206	COPD
IL1RN	ENSG00000136689	rs2234663	COPD
IL1RN	ENSG00000136689	rs2234663	COPD
IREB2	ENSG00000136381	rs11858836	COPD
IREB2	ENSG00000136381	rs13180	COPD
IREB2	ENSG00000136381	rs2568494	COPD
IREB2	ENSG00000136381	rs2568494	COPD
IREB2	ENSG00000136381	rs2568494	COPD
IREB2 (MMA)	ENSG00000136381	rs2568494	COPD
IREB2	ENSG00000136381	rs2656052	COPD
IREB2	ENSG00000136381	rs2656052	COPD
IREB2	ENSG00000136381	rs2656065	COPD
IREB2	ENSG00000136381	rs2656065	COPD
IREB2	ENSG00000136381	rs2938670	COPD
IREB2	ENSG00000136381	rs2938670	COPD
KBTBD12	ENSG00000187715	rs17282209	COPD
KBTBD12	ENSG00000187715	rs17282209	COPD
MICAL1	ENSG00000135596	rs59056467	COPD
MMP3	ENSG00000149968	rs679620	COPD
MMP9	ENSG00000100985	rs17576	COPD
MMP9	ENSG00000100985	rs17576	COPD
MMP9	ENSG00000100985	rs3918242	COPD
MMP12	ENSG00000262406	rs626750	Severe COPD
MTCL1	ENSG00000168502	rs647097	COPD
PID1	ENSG00000153823	rs16825267	COPD
PSMA4	ENSG00000041357	rs58365910	COPD
PSMA4	ENSG00000041357	rs58365910	COPD
RAB4B	ENSG00000167578	rs2604894	COPD
RAB4B	ENSG00000167578	rs7937	COPD
RARB	ENSG00000077092	rs1529672	COPD
RIN3	ENSG00000100599	rs1075472	COPD
RIN3	ENSG00000100599	rs1075472	COPD
RIN3	ENSG00000100599	rs72699855	COPD
RIN3	ENSG00000100599	rs754388	COPD
RIN3	ENSG00000100599	rs754388	COPD
RIN3	ENSG00000100599	rs754388	COPD

SERPINA1	ENSG00000197249	PiSZ	COPD
SERPINA1	ENSG00000197249	PiMZ	COPD
SERPINA1	ENSG00000197249	rs28929474	COPD
SFTPA1/SFTPA2/SFTPB/SFTPD combined	ENSG00000122852/ ENSG00000185303/ ENSG00000133661	Combination of rs1059046, rs1136451, rs4253527, rs1130866, rs2077079, rs1051246, rs2245121, rs2255601, rs3088308, rs6413520, rs721917 and rs911887	COPD
SFTPA1/SFTPA2/SFTPB/SFTPD combined	ENSG00000122852/ ENSG00000185303/ /ENSG00000133661	Combination of rs1059046, rs1136451, rs4253527, rs1130866, rs2077079, rs1051246, rs2245121, rs2255601, rs3088308, rs6413520, rs721917 and rs911887	COPD
SFTPA1/SFTPA2	ENSG00000122852/ ENSG00000185303	Combination of rs1059046, rs1136451 and rs4253527	COPD
SFTPD	ENSG00000133661	rs721917	COPD
TET2	ENSG00000168769	rs2047409	COPD
TGFB1	ENSG00000105329	rs1800470	COPD
TGFB2	ENSG00000092969	rs4846480	Severe COPD
TGFB2	ENSG00000092969	rs10429950	COPD
TIRAP	ENSG00000150455	rs8177374	COPD
THSD4	ENSG00000187720	rs1441358	COPD
TNFA	ENSG00000232810	rs1800629	COPD
TNFA	ENSG00000232810	rs1800629	COPD
TNFA	ENSG00000232810	rs1800630	COPD
TNFA	ENSG00000232810	rs1800630	COPD
TNFA	ENSG00000232810	rs80267959	COPD
TNFA	ENSG00000232810	rs80267959	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VDBP	ENSG00000145321	rs4588, rs7041	COPD
VPS53	ENSG00000141252	rs11247558	COPD
VPS53	ENSG00000141252	rs11656538	COPD
VPS53	ENSG00000141252	rs34001232	COPD
VPS53	ENSG00000141252	rs34469205	COPD
VPS53	ENSG00000141252	rs34729304	COPD

VPS53	ENSG00000141252	rs35716682	COPD
VPS53	ENSG00000141252	rs4968100	COPD

List of SNPs and genes selected from the comprehensive overview.

Table E2

Gene identifier	Gene name	Full gene name	Synthetic function	General classification	Functional network area	GeneCards	Date accessed	Sources
ENSG00000159640	ACE	Angiotensin I Converting Enzyme	Homeostatic control	Blood pressure	Homeostasis-organismal	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ACE&keywords=ace	20-4-2018	E(7, 8)
ENSG00000135074	ADAM19	Disintegrin And Metalloproteinase Domain-Containing Protein	Cellular interactions	Transmembrane protein	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAM19&keywords=ADAM19	20-4-2018	E(9-11)
ENSG00000149451	ADAM33	Disintegrin And Metalloproteinase Domain 33	Cellular interactions	Transmembrane protein	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAM33&keywords=ADAM33	20-4-2018	E(12-15)
ENSG00000112414	ADGRG6	Adhesion G Protein-Coupled Receptor G6	Cellular interactions/ signalling	Transmembrane protein	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADGRG6&keywords=ADGRG6	20-4-2018	E(16-18)
ENSG00000169252	ADRB2	Adrenoceptor Beta 2 Surface	Cellular interactions/ signalling	Transmembrane protein	Homeostasis-organismal	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADRB2&keywords=ADRB2	20-4-2018	E(19)
ENSG00000204305	AGER	Advanced Glycation End-Product Specific Receptor	Cellular interactions/ inflammation	Transmembrane protein	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=AGER&keywords=AGER	20-4-2018	E(20)
ENSG00000155085	AK9	Adenylate Kinase 9	Cellular metabolism	Kinase	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=AK9&keywords=AK9	20-4-2018	E(21, 22)
ENSG00000118690	ARMC2	Armadillo Repeat Containing 2	Protein coding	Protein coding	Cellular-metabolism*	https://amp.pharm.mssm.edu/archs4/search/genepage.php?search=go&gene=ARMC2	20-4-2018	E(23, 24)

ENSG00000075275	CELSR1	Cadherin EGF LAG Seven-Pass G-Type Receptor 1	Cellular interactions/ signalling	Cadherins	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CELSR1&keywords=CELSR1	20-4-2018	E(25)
ENSG00000153774	CFDP1	Craniofacial Development Protein 1	Protein coding	Protein coding	Cellular-metabolism*	https://amp.pharm.mssm.edu/archs4/search/genepage.php?search=go&gene=CFDP1	20-4-2018	E(23)
ENSG00000080644	CHRNA3	Cholinergic Receptor Nicotinic Alpha 3 Subunit	Neurotransmission	Ion-channel	Addiction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHRNA3&keywords=CHRNA3	20-4-2018	E(26, 27)
ENSG00000169684	CHRNA5	Cholinergic Receptor Nicotinic Alpha 5 Subunit	Neurotransmission	Ion-channel	Addiction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHRNA5&keywords=CHRNA5	20-4-2018	E(28)
ENSG00000117971	CHRNB4	Cholinergic Receptor Nicotinic Beta 4 Subunit	Neurotransmission	Ion-channel	Addiction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHRNB4&keywords=CHRNB4	20-4-2018	E(27)
ENSG00000007545	CRAMP1L	Cramped Chromatin Regulator Homolog 1	Protein coding	Protein coding	Unknown	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CRAMP1&keywords=CRAMP1L	20-4-2018	E(29)
ENSG00000140465	CYP1A1	Cytochrome P450 Family 1 Subfamily A Member 1	Metabolism	Drug metabolism and cholesterol synthesis	Detoxification	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CYP1A1&keywords=CYP1A1	20-4-2018	E(30-32).
ENSG00000255974	CYP2A6	Cytochrome P450 Family 2 Subfamily A Member 6	Metabolism	Drug metabolism and cholesterol synthesis	Detoxification	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CYP2A6&keywords=CYP2A6	20-4-2018	E(33, 34)
ENSG00000096696	DSP	Desmoplakin	Structure	Structural protein	Cellular-structure	https://www.genecards.org/cgi-bin/carddisp.pl?gene=DSP&keywords=DSP	20-4-2018	E(35)
ENSG00000132394	EEFSEC	Eukaryotic Elongation Factor, Selenocysteine-TRNA Specific	Protein Coding	Protein coding	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=EEFSEC&keywords=EEFSEC	20-4-2018	E(36, 37)

ENSG00000224032	EPB41L4A-AS1	(EPB41L4A Antisense RNA 1	Antisense RNA	RNA coding	Unknown	https://www.genecards.org/cgi-bin/carddisp.pl?gene=EPB41L4A-AS1&keywords=EPB41L4A-AS1	20-4-2018	E(38)
ENSG00000143819	EPHX1	Epoxide Hydrolase 1	Metabolism	Epoxide metabolism	Detoxification	https://www.genecards.org/cgi-bin/carddisp.pl?gene=EPHX1&keywords=EPHX1	20-4-2018	E(39, 40)
ENSG00000138640	FAM13A	Family With Sequence Similarity 13 Member A	Protein coding	Protein coding	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=FAM13A&keywords=FAM13A	20-4-2018	(41, 42)
ENSG00000108021	FAM208B	Family With Sequence Similarity 208 Member B	Protein coding	Protein coding	Cellular-structure*	https://amp.pharm.mssm.edu/archs4/search/genepage.php?search=go&gene=FAM208B	20-4-2018	Not available
ENSG00000064763	FAR2	Fatty Acyl-CoA Reductase 2	Metabolism	Wax biosynthesis	Metabolism-organismal	https://www.genecards.org/cgi-bin/carddisp.pl?gene=FAR2&keywords=FAR2	20-4-2018	EE(43)
ENSG00000150907	FOXO1	Forkhead Box O1	Metabolism	Transcription factor that is the main target of insulin signaling	Regulation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOXO1&keywords=ENSG00000150907	20-4-2018	E(44, 45)
ENSG00000179409	GEMIN4	Gem Nuclear Organelle Associated Protein 4	Cellular metabolism	Spliceosome regeneration	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=GEMIN4&keywords=GEMIN4	20-4-2018	E(46).
ENSG00000138780	GSTCD	Glutathione S-Transferase C-Terminal Domain Containing	Protein coding	Protein Coding	Detoxification	https://www.genecards.org/cgi-bin/carddisp.pl?gene=GSTCD&keywords=GSTCD	20-4-2018	E(47)
ENSG00000134184	GSTM1	glutathione S-transferase mu 1	Metabolism	Drug metabolism	Detoxification	https://www.genecards.org/cgi-bin/carddisp.pl?gene=GSTM1&keywords=GSTM1	20-4-2018	E(48, 49)
ENSG00000170180	GYPA	Glycophorin A (MNS Blood Group)	Structure	Structural protein	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=GYPA&keywords=E	20-4-2018	E(50).

						NSG00000170180		
ENSG00000164161	HHIP	Hedgehog Interacting Protein	Structure	Embryonic development	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HHIP&keywords=HHIP	20-4-2018	E(51)
ENSG00000100292	HMOX1	Heme Oxygenase 1	Metabolism	Endogenous metabolism	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HMOX1&keywords=HMOX1	20-4-2018	E(52)
ENSG00000164270	HTR4	5-Hydroxytryptamine Receptor 4	Neurotransmission	NT receptor	Addiction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HTR4&keywords=ENSG00000164270	20-4-2018	E(53)
ENSG00000169194	IL13	Interleukin 13	Anti-inflammatory	Anti-inflammatory cytokine	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=IL13&keywords=ENSG00000169194	20-4-2018	E(54).
ENSG00000125538	IL1B	Interleukin 1 Beta	Inflammation	Inflammatory cytokine	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=IL1B&keywords=ENSG00000125538	20-4-2018	E(54)
ENSG00000136689	IL1RN	Interleukin 1 Receptor Antagonist	Anti-inflammatory	Inflammation regulation	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=IL1RN&keywords=ENSG00000136689	20-4-2018	E(54)
ENSG00000197272	IL27	Interleukin 27	Inflammation	Inflammatory/ anti inflammatory cytokine	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=IL27&keywords=ENSG00000197272	20-4-2018	E(54)
ENSG00000136244	IL6	Interleukin 6	Inflammation	Inflammatory cytokine	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=IL6&keywords=ENSG00000136244	20-4-2018	E(54)

ENSG00000136381	IREB2	Iron Responsive Element Binding Protein 2	Homeostatic control	Iron metabolism	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=IREB2&keywords=ENSG00000136381	20-4-2018	E(55)
ENSG00000187715	KBTBD12	Kelch Repeat And BTB Domain Containing 12	Protein coding	Protein coding	Cellular-metabolism*	http://www.informatics.jax.org/marker/MGI:1918481	20-4-2018	Not available
ENSG00000135596	MICAL1	Microtubule Associated Monooxygenase, Calponin And LIM Domain Containing 1	Structure	Protein mobilisation?	Cellular-structure	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MICAL1&keywords=MICAL1	20-4-2018	E(56)
ENSG00000262406	MMP12	Matrix Metallopeptidase 12	Metalloproteinase	Cellular degradation	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MMP12&keywords=MMP12	20-4-2018	E(57, 58)
ENSG00000149968	MMP3	Matrix Metallopeptidase 3	Metalloproteinase	Cellular degradation	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MMP3&keywords=MMP3	20-4-2018	E(57, 58)
ENSG00000100985	MMP9	Matrix Metallopeptidase 9	Metalloproteinase	Cellular degradation	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MMP9&keywords=MMP9	20-4-2018	E(57, 58)
ENSG00000204472	AIF1	Allograft Inflammatory Factor 1	Protein coding	Macrophage activation	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=AIF1&keywords=AIF1	21-7-2018	E(59)
ENSG00000168502	MTCL1	Microtubule Cross Linking Factor 1	Protein coding	Protein Coding	Cellular-structure	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MTCL1&keywords=MTCL1	22-4-2018	E(60, 61)
ENSG00000153823	PID1	phosphotyrosine Interaction Domain Containing 1	Protein Coding	Protein Coding	Regulation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=PID1&keywords=ENSG00000153823	22-4-2018	(62-64)

ENSG00000041357	PSMA4	Proteasome Subunit Alpha 4	Cellular metabolism	Protein degradation	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=PSMA4&keywords=ENSG00000041357	22-4-2018	(65)
ENSG00000167578	RAB4B	RAB4B, Member RAS Oncogene Family	Cellular metabolism	Protein degradation?	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=RAB4B&keywords=ENSG00000167578	22-4-2018	(66)
ENSG00000077092	RARB	Retinoic Acid Receptor Beta	Transcription factor	Growth regulation	Regulation (transcription factor	https://www.genecards.org/cgi-bin/carddisp.pl?gene=RARB&keywords=ENSG00000077092	22-4-2018	EE(59)
ENSG00000185305	ARL15	ADP Ribosylation Factor Like GTPase 15 2 3 5	Protein coding	GTP binding	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ARL15	22-7-2018	(67, 68) (69)
ENSG00000100599	RIN3	Ras And Rab Interactor 3	Cellular metabolism	Intracellular membrane trafficking,	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=RIN3&keywords=ENSG00000100599	23-4-2018	E(70, 71)
ENSG00000197249	SERPINA1	Serpin Family A Member 1	Structure	Structural protein	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SERPINA1&keywords=ENSG00000197249	23-4-2018	E(62-64)
ENSG00000122852	SFTPA1	Surfactant Protein A1	Immunology/ Lung function	Immune function/lung function	Lung function	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SFTPA1&keywords=SFTPA1	23-4-2018	E(72)
ENSG00000185303	SFTPA2	Surfactant Protein A2	Lung function	Lung function	Lung function	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SFTPA2&keywords=SFTPA2	23-4-2018	E(72)
ENSG00000133661	SFTPD	Surfactant Protein D	Immunology/ Lung function	Immune function	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SFTPD&keywords=SFTPD	23-4-2018	E(73, 74)
ENSG00000176476	SGF29	SAGA Complex Associated Factor 29	Transcription	DNA transcription	Regulation	http://www.genecards.org/cgi-bin/carddisp.pl?gene=SGF29	23-4-2018	E(75)

						&keywords=SGF29		
ENSG00000105329	TGFB1	Transforming Growth Factor Beta 1	Regulator	Master regulator	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TGFB1&keywords=TGFB1	23-4-2018	E(76)
ENSG00000092969	TGFB2	Transforming Growth Factor Beta 2	Regulator	Master regulator	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TGFB2&keywords=TGFB2	23-4-2018	E(77-79)
ENSG00000187720	THSD4	Thrombospondin Type 1 Domain Containing 4	Regulator	Regulator	Regulation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=THSD4&keywords=THSD4	23-4-2018	E(80-82)
ENSG00000150455	TIRAP	TIR Domain Containing Adaptor Protein	Immunology	Immune function/ inflammation	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TIRAP&keywords=TIRAP	23-4-2018	E(83, 84)
ENSG00000232810	TNFA	Tumor Necrosis Factor	Inflammation	Inflammation regulation	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TNF&keywords=TNFA	23-4-2018	E(85)
ENSG00000145321	VDBP	GC, Vitamin D Binding Protein	Metabolism	Vitamin D transport	Homeostasis-organismal	https://www.genecards.org/cgi-bin/carddisp.pl?gene=GC	23-4-2018	E(86, 87)
ENSG00000141252	VPS53	VPS53, GARP Complex Subunit	Cellular metabolism	Endosome recycling	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=VPS53	23-4-2018	E(88)
ENSG00000148219	ASTN2	Astrotactin 2	Protein coding	Glial-guided neuronal migration	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ASTN2&keywords=ASTN2	23-7-2018	E(89)
ENSG00000197536	C5orf56	Chromosome 5 Open Reading Frame 56	Protein coding	Protein coding	Unknown	https://www.genecards.org/cgi-bin/carddisp.pl?gene=C5orf56&keywords=C5orf56	23-7-2018	Not available.
ENSG00000157445	CACNA2D3	Calcium Voltage-Gated Channel Auxiliary Subunit	Protein coding	Voltage gated signal transduction	Addiction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CACNA2D3&keywo	23-7-2018	E(90, 91)

		Alpha2delta 3				rds=CACNA2D3		
ENSG00000151465	CDC123	Cell Division Cycle 123	Protein coding	Cellular stress response	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CDC123&keywords=CDC123	23-7-2018	E(92-94)
ENSG00000055163	CYFIP2	Cytoplasmic FMR1 Interacting Protein	Protein coding	Axon sorting/ neural development	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CYFIP2&keywords=CYFIP2	23-7-2018	(95)
ENSG00000107249	GLIS3	GLIS Family Zinc Finger 3	Protein coding	Regulation of transcription	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=GLIS3&keywords=GLIS3	23-7-2018	E(92-94)
ENSG00000068024	HDAC4	Histone Deacetylase 4	Protein coding	Regulation of transcription	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HDAC4	23-7-2018	E(96, 97)
ENSG00000237541	HLA-DQA2	Major Histocompatibility Complex, Class II, DQ Alpha 2	Antigen presenting receptor	Immune response/protein binding/ receptor	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HLA-DQA2&keywords=HLA-DQA2	23-7-2018	E(98)
ENSG00000179344	HLA-DQB1	Major Histocompatibility Complex, Class II, DQ Beta 1	Antigen presenting receptor	Immune response/protein binding/ receptor	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HLA-DQB1	23-7-2018	E(98)
ENSG00000188266	HYKK	Hydroxylysine Kinase	Protein coding	Kinase	Addiction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HYKK	23-7-2018	E(99-101)
ENSG00000213949	ITGA1	Integrin Subunit Alpha 1	Protein coding	Cellular interactions and adhesion	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=ITGA1	23-7-2018	E(102)
ENSG00000159197	KCNE2	Potassium Voltage-Gated Channel Subfamily E Regulatory Subunit 2	Signal transmission	Voltage-gated ion channels	Cellular-interaction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNE2	23-7-2018	E(103)
ENSG00000185760	KCNQ5	Potassium Voltage-Gated Channel Subfamily Q Member 5	Signal transmission	Voltage-gated ion channels	Cellular-interaction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNQ5&keywords=	23-7-2018	E(104, 105)

						ENSG00000185760		
ENSG00000170745	KCNS3	Potassium Voltage-Gated Channel Modifier Subfamily S Member 3	Signal transmission	Voltage-gated ion channels	Cellular-interaction	https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNS3	23-7-2018	E(106, 107)
ENSG00000227456	LINC00310	Long Intergenic Non-Protein Coding RNA 310	RNA gene	Unknown	Unknown	https://www.genecards.org/cgi-bin/carddisp.pl?gene=LINC00310	23-7-2018	E(108)
ENSG00000148655	LRMDA	Leucine Rich Melanocyte Differentiation Associated	Protein coding	Melanocyte differentiation	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=LRMDA&keywords=LRMDA	23-7-2018	Not available
ENSG00000123384	LRP1	LDL Receptor Related Protein	Transmembrane receptor	LDL receptor protein	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=LRP1&keywords=LRP1	23-7-2018	E(109-111)
ENSG00000085276	MECOM	MDS1 And EVI1 Complex Locus	Protein coding	Transcription regulator	Regulation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MECOM&keywords=MECOM	23-7-2018	E(112)
ENSG00000117122	MFAP2	Microfibril Associated Protein 2	Structure	Major antigen of elastin-associated microfibrils	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MFAP2&keywords=MFAP2	23-7-2018	E(113)
ENSG00000102996	MMP15	Matrix Metallopeptidase 15	Metalloproteinase	Cellular degradation	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MMP15&keywords=ENSG00000102996	23-7-2018	E(57, 58)
ENSG00000169184	MN1	MN1 Proto-Oncogene, Transcriptional Regulator	Transcription	Transcription regulator, oncogene	Regulation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MN1&keywords=MN1	23-7-2018	E(114, 115)
ENSG00000204475	NCR3	natural Cytotoxicity Triggering Receptor 3	Receptor	Natural killer cell interaction	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=NCR3&keywords=NCR3	23-7-2018	E(116)

ENSG00000168743	NPNT	Nephronectin	Ligand	Ligand of integrin alpha-8/beta	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=NPNT&keywords=NPNT	23-7-2018	E(117-119)
ENSG00000090621	PABPC4	Poly(A) Binding Protein Cytoplasmic 4	RNA protein	RNA-processing protein	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=PABPC4&keywords=PABPC4#proteins	23-7-2018	E(120)
ENSG00000019485	PRDM11	PR/SET Domain 11	Protein coding	Transcriptional regulator	Regulation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=PRDM11&keywords=PRDM11	23-7-2018	E(121, 122).
ENSG00000185920	PTCH1	Patched 1	Protein coding	Related to SHH genes and tumour suppressors	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCH1&keywords=PTCH1	23-7-2018	E(123, 124)
ENSG00000240857	RDH14	Retinol Dehydrogenase 14	Enzyme	Purine metabolism	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=RDH14&keywords=RDH14	23-7-2018	E(125)
ENSG00000168065	SLC22A11	Solute Carrier Family 22 Member 11	Transport protein	Salt metabolism/excretion	Metabolism-organismal	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC22A11&keywords=SLC22A11	23-7-2018	E(126)
ENSG00000139343	SNRPF	Small Nuclear Ribonucleoprotein Polypeptide F	Protein coding	RNA-processing protein	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SNRPF&keywords=SNRPF	23-7-2018	Not available
ENSG00000145757	SPATA9	Spermatogenesis Associated 9	Protein coding	Unknown	Development	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SPATA9&keywords=SPATA9	23-7-2018	E(127)
ENSG00000153820	SPHKAP	SPHK1 Interactor, AKAP Domain Containing	Enzyme	Kinase	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=SPHKAP&keywords=SPHKAP	23-7-2018	E(128)
ENSG00000153060	TEKT5	Tektin 5	Protein coding	Unknown	Cellular-structure	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TEKT5&keywords=TEKT5	23-7-2018	E(129)

ENSG00000132749	TESMIN *	Testis Expressed Metallothionein Like Protein	heavy metal binding	Metallothionein protein	Cellular-metabolism	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TESMIN	23-7-2018	E(130)
ENSG00000168769	TET2	Tet Methylcytosine Dioxygenase 2	Enzyme	Methylcytosine dioxygenase	Inflammation	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TET2	23-7-2018	E(131, 132)
ENSG00000079308	TNS1	Tensin 1	Protein coding	Extracellular matrix cross linking	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=TNS1&keywords=TNS1	23-7-2018	E(85)
ENSG00000100815	TRIP11	Thyroid Hormone Receptor Interactor 11	Receptor protein in golgi	Transcription-coregulator	Regulation	https://www.genecards.org/Search/Keyword?queryString=TRIP11	23-7-2018	E(133)
ENSG00000114251	WNT5A	Wnt Family Member 5A	Ligand for transmembrane proteins	Ligand	Tissue-remodelling	https://www.genecards.org/cgi-bin/carddisp.pl?gene=WNT5A&keywords=WNT5A	23-7-2018	E(134)

Gene descriptions with functional classes.

Supplementary references

E1. Zambon AC, Gaj S, Ho I, Hanspers K, Vranizan K, Evelo CT, Conklin BR, Pico AR, Salomonis N. GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* 2012; 28: 2209-2210.

E2. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. *Genome biology* 2016; 17: 122.

E3. Bruce Alberts A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell, ed. Garland Science. Taylor & Francis Group, New York; 2008.

E4. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nature protocols* 2016; 11: 1.

E5. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome research* 2001; 11: 863-874.

E6. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics* 2013; 76: 7.20. 21-27.20. 41.

E7. Tikhomirova VE, Kost OA, Kryukova OV, Golukhova EZ, Bulaeva NI, Zholbaeva AZ, Bokeria LA, Garcia JG, Danilov SM. ACE phenotyping in human heart. *PloS one* 2017; 12: e0181976.

E8. Gonzalez-Villalobos RA, Shen XZ, Bernstein EA, Janjulia T, Taylor B, Giani JF, Blackwell W-LB, Shah KH, Shi PD, Fuchs S. Rediscovering ACE: novel insights into the many roles of the angiotensin-converting enzyme. *Journal of molecular medicine* 2013; 91: 1143-1154.

E9. Pérez-Rubio G, Silva-Zolezzi I, Fernández-López JC, Camarena Á, Velázquez-Uncal M, Morales-Mandujano F, Hernández-Zenteno RDJ, Flores-Trujillo F, Sánchez-Romero C, Velázquez-Montero A. Genetic variants in IL6R and ADAM19 are associated with COPD severity in a Mexican Mestizo population. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2016; 13: 610-615.

E10. Zhang Q, Yu L, Qin D, Huang R, Jiang X, Zou C, Tang Q, Chen Y, Wang G, Wang X. Role of microRNA-30c targeting ADAM19 in colorectal cancer. *PLoS One* 2015; 10: e0120698.

E11. London SJ, Gao W, Gharib SA, Hancock DB, Wilk JB, House JS, Gibbs RA, Muzny DM, Lumley T, Franceschini N. ADAM19 and HTR4 variants and pulmonary function: cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium targeted sequencing study. *Circulation: Genomic and Precision Medicine* 2014; 7: 350-358.

E12. Wang X, Li L, Xiao J, Jin C, Huang K, Kang X, Wu X, Lv F. Association of ADAM33 gene polymorphisms with COPD in a northeastern Chinese population. *BMC medical genetics* 2009; 10: 132.

E13. Weerasekera L, Rudnicka C, Sang Q-X, Curran JE, Johnson MP, Moses EK, Göring HH, Blangero J, Hricova J, Schlaich M. ADAM19: A novel target for metabolic syndrome in humans and mice. *Mediators of Inflammation* 2017; 2017.

E14. Shapiro SD, Owen CA. ADAM-33 surfaces as an asthma gene. *New England Journal of Medicine* 2002; 347: 936-938.

E15. Aierken H, Wang J, Wushouer Q, Shayhidin E, Hu X, Syed I, Wufuer D. Polymorphisms of the ADAM 33 gene and chronic obstructive pulmonary disease risk: a meta-analysis. *The clinical respiratory journal* 2014; 8: 108-115.

E16. Mogha A, Harty BL, Carlin D, Joseph J, Sanchez NE, Suter U, Piao X, Cavalli V, Monk KR. Gpr126/Adgrg6 has Schwann cell autonomous and nonautonomous functions in peripheral nerve injury and repair. *Journal of Neuroscience* 2016; 36: 12351-12367.

E17. Karner CM, Long F, Solnica-Krezel L, Monk KR, Gray RS. Gpr126/Adgrg6 deletion in cartilage models idiopathic scoliosis and pectus excavatum in mice. *Human molecular genetics* 2015; 24: 4365-4373.

E18. Cazorla-Vázquez S, Engel FB. Adhesion GPCRs in Kidney Development and Disease. *Frontiers in Cell and Developmental Biology* 2018; 6: 9.

E19. Nielsen AO, Jensen CS, Arredouani MS, Dahl R, Dahl M. Variants of the ADRB2 gene in COPD: systematic review and meta-analyses of disease risk and treatment response. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2017; 14: 451-460.

E20. Yin N, Lang X, Wang X, Liu W. AGER genetic polymorphisms increase risks of breast and lung cancers. *Genet Mol Res* 2015; 14: 17776-17787.

E21. Edwards L, Gupta R, Filipp FV. Hypermutation of DPYD deregulates pyrimidine metabolism and promotes malignant progression. *Molecular Cancer Research* 2016; 14: 196-206.

E22. Lam C-W, Wong K-S, Leung H-W, Law C-Y. Limb girdle myasthenia with digenic RAPSN and a novel disease gene AK9 mutations. *European Journal of Human Genetics* 2017; 25: 192.

E23. Artigas MS, Loth DW, Wain LV, Gharib SA, Obeidat Me, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature genetics* 2011; 43: 1082.

E24. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC, Ma'ayan A. Massive mining of publicly available RNA-seq data from human and mouse. *Nature communications* 2018; 9: 1366.

E25. Hardin M, Cho MH, Sharma S, Glass K, Castaldi PJ, McDonald M-L, Aschard H, Senter-Sylvia J, Tantisira K, Weiss ST. Sex-based genetic association study identifies CELSR1 as a possible chronic obstructive pulmonary disease risk locus among women. *American journal of respiratory cell and molecular biology* 2017; 56: 332-341.

E26. Wu X-Y, Zhou S-Y, Niu Z-Z, Liu T, Xie C-B, Chen W-Q. CHRNA3 rs6495308 Genotype as an Effect Modifier of the Association between Daily Cigarette Consumption and Hypertension in Chinese Male Smokers. *International journal of environmental research and public health* 2015; 12: 4156-4169.

E27. Shmulewitz D, Meyers JL, Wall MM, Aharonovich E, Frisch A, Spivak B, Weizman A, Edenberg HJ, Gelernter J, Hasin DS. CHRNA5/A3/B4 Variant rs3743078 and Nicotine-

Related Phenotypes: Indirect Effects Through Nicotine Craving. *Journal of studies on alcohol and drugs* 2016; 77: 227-237.

E28. Wang Q, Li S, Li H, Yang X, Jiang F, Zhang N, Han M, Jia C. Relative contribution of genetic, individual, and self-efficacy factors to smoking cessation in a Chinese rural population. *The American journal on addictions* 2017; 26: 161-166.

E29. CRAMP1 cramped chromatin regulator homolog 1 [Homo sapiens (human)]. [Database] 6-1-2019 [cited 2018. Available from: https://www.ncbi.nlm.nih.gov/gene/57585.

E30. Vibhuti A, Arif E, Mishra A, Deepak D, Singh B, Rahman I, Mohammad G, Pasha MAQ. CYP1A1, CYP1A2 and CYBA gene polymorphisms associated with oxidative stress in COPD. *Clin Chim Acta* 2010; 411: 474-480.

E31. Wang C-D, Chen N, Huang L, Wang J-R, Chen Z-Y, Jiang Y-M, He Y-Z, Ji Y-L. Impact of CYP1A1 polymorphisms on susceptibility to chronic obstructive pulmonary disease: a meta-analysis. *BioMed research international* 2015; 2015.

E32. Badal S, Delgoda R. Role of the modulation of CYP1A1 expression and activity in chemoprevention. *Journal of Applied Toxicology* 2014; 34: 743-753.

E33. Akrodou YM. CYP2A6 Polymorphisms May Strengthen Individualized Treatment for Nicotine Dependence. *Scientifica* 2015; 2015: 7.

E34. Tanner J-A, Prasad B, Claw KG, Stapleton P, Chaudhry A, Schuetz EG, Thummel KE, Tyndale RF. Predictors of Variation in CYP2A6 mRNA, Protein, and Enzyme Activity in a Human Liver Bank: Influence of Genetic and Nongenetic Factors. *The Journal of pharmacology and experimental therapeutics* 2017; 360: 129-139.

E35. Vasioukhin V, Bowers E, Bauer C, Degenstein L, Fuchs E. Desmoplakin is essential in epidermal sheet formation. *Nature Cell Biology* 2001; 3: 1076.

E36. Sakornsakolpat P, Morrow JD, Castaldi PJ, Hersh CP, Bossé Y, Silverman EK, Manichaikul A, Cho MH. Integrative genomics identifies new genes associated with severe COPD and emphysema. *Respiratory research* 2018; 19: 46-46.

E37. Zhang G, Feenstra B, Bacelis J, Liu X, Muglia LM, Juodakis J, Miller DE, Litterman N, Jiang P-P, Russell L, Hinds DA, Hu Y, Weirauch MT, Chen X, Chavan AR, Wagner GP, Pavličev M, Nnamani MC, Maziarz J, Karjalainen MK, Rämet M, Sengpiel V, Geller F, Boyd HA, Palotie A, Momany A, Bedell B, Ryckman KK, Huusko JM, Forney CR, Kottyan LC, Hallman M, Teramo K, Nohr EA, Davey Smith G, Melbye M, Jacobsson B, Muglia LJ. Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *The New England journal of medicine* 2017; 377: 1156-1167.

E38. EPB41L4A antisense RNA 1 [Homo sapiens (human)]. 5-8-2018 [cited 2018. Available from: https://www.ncbi.nlm.nih.gov/gene/114915EPB41L4A-AS1

E39. Václavíková R, Hughes DJ, Souček P. Microsomal epoxide hydrolase 1 (EPHX1): Gene, structure, function, and role in human disease. *Gene* 2015; 571: 1-8.

E40. Chen H, Ge L, Sui Q, Lin M. Systematic Review and Meta-Analysis of the Relationship between EPHX1 Polymorphisms and the Risk of Head and Neck Cancer. *PloS one* 2015; 10: e0123347-e0123347.

E41. Ziółkowska-Suchanek I, Mosor M, Gabryel P, Grabicki M, Żurawek M, Fichna M, Strauss E, Batura-Gabryel H, Dyszkiewicz W, Nowak J. Susceptibility loci in lung cancer and COPD: association of IREB2 and FAM13A with pulmonary diseases. *Scientific Reports* 2015; 5: 13502.

E42. Hawkins GA, Mora AL. FAM13A, A Fatty Acid Oxidation Switch in Mitochondria. Friend or Foe in Chronic Obstructive Pulmonary Disease Pathogenesis? *American journal of respiratory cell and molecular biology* 2017; 56: 689-691.

E43. Honsho M, Asaoku S, Fukumoto K, Fujiki Y. Topogenesis and homeostasis of fatty acyl-CoA reductase 1. *The Journal of biological chemistry* 2013; 288: 34588-34598.

E44. Huang H, Tindall DJ. Dynamic FoxO transcription factors. *J Cell Sci* 2007; 120: 2479-2487.

E45. Barthel A, Schmoll D, Unterman TG. FoxO proteins in insulin action and metabolism. *Trends in Endocrinology & Metabolism* 2005; 16: 183-189.

E46. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *cell* 2004; 116: 281-297.

E47. Soler Artigas M, Wain LV, Repapi E, Obeidat Me, Sayers I, Burton PR, Johnson T, Zhao JH, Albrecht E, Dominiczak AF. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. *American journal of respiratory and critical care medicine* 2011; 184: 786-795.

E48. Khalighinasab MR, Saify K, Saadat M. Association between GSTM1 and GSTT1 polymorphisms and susceptibility to methamphetamine dependence. *Molecular biology research communications* 2015; 4: 25-32.

E49. Krüger M, Pabst AM, Mahmoodi B, Becker B, Kämmerer PW, Koch FP. The impact of GSTM1/GSTT1 polymorphism for the risk of oral cancer. *Clinical Oral Investigations* 2015; 19: 1791-1797.

E50. Bigham AW, Magnaye K, Dunn DM, Weiss RB, Bamshad M. Complex signatures of natural selection at GYPA. *Human Genetics* 2018; 137: 151-160.

E51. Wang B, Zhou H, Yang J, Xiao J, Liang B, Li D, Zhou H, Zeng Q, Fang C, Rao Z, Yu H, Ou X, Feng Y. Association of HHIP polymorphisms with COPD and COPD-related phenotypes in a Chinese Han population. *Gene* 2013; 531: 101-105.

E52. Zhou H, Ying X, Liu Y, Ye S, Yan J, Li Y. Genetic polymorphism of heme oxygenase 1 promoter in the occurrence and severity of chronic obstructive pulmonary disease: a meta-analysis. *Journal of cellular and molecular medicine* 2017; 21: 894-903.

E53. Bockaert J, Claeysen S, Compan V, Dumuis A. 5-HT4 receptors: History, molecular pharmacology and brain functions. *Neuropharmacology* 2008; 55: 922-931.

E54. Akdis M, Burgler S, Crameri R, Eiwegger T, Fujita H, Gomez E, Klunker S, Meyer N, O'Mahony L, Palomares O, Rhyner C, Quaked N, Schaffartzik A, Van De Veen W, Zeller S, Zimmermann M, Akdis CA. Interleukins, from 1 to 37, and interferon-γ: Receptors, functions, and roles in diseases. *Journal of Allergy and Clinical Immunology* 2011; 127: 701-721.e770.

E55. Kim WJ, Lee SD. Candidate genes for COPD: current evidence and research. *International journal of chronic obstructive pulmonary disease* 2015; 10: 2249-2255.

E56. Deng W, Wang Y, Zhao S, Zhang Y, Chen Y, Zhao X, Liu L, Sun S, Zhang L, Ye B, Du J. MICAL1 facilitates breast cancer cell proliferation via ROS-sensitive ERK/cyclin D pathway. *Journal of cellular and molecular medicine* 2018; 22: 3108-3118.

E57. Misra S, Talwar P, Kumar A, Kumar P, Sagar R, Vibha D, Pandit AK, Gulati A, Kushwaha S, Prasad K. Association between matrix metalloproteinase family gene polymorphisms and risk of ischemic stroke: A systematic review and meta-analysis of 29 studies. *Gene* 2018; 672: 180-194.

E58. Ye S. Polymorphism in matrix metalloproteinase gene promoters: implication in regulation of gene expression and susceptibility of various diseases. *Matrix Biology* 2000; 19: 623-629.

E59. Zhao Y-Y, Yan D-J, Chen Z-W. Role of AIF-1 in the regulation of inflammatory activation and diverse disease processes. *Cellular immunology* 2013; 284: 75-83.

E60. Sato Y, Hayashi K, Amano Y, Takahashi M, Yonemura S, Hayashi I, Hirose H, Ohno S, Suzuki A. MTCL1 crosslinks and stabilizes non-centrosomal microtubules on the Golgi membrane. *Nature Communications* 2014; 5: 5266.

E61. Satake T, Yamashita K, Hayashi K, Miyatake S, Tamura-Nakano M, Doi H, Furuta Y, Shioi G, Miura E, Takeo YH, Yoshida K, Yahikozawa H, Matsumoto N, Yuzaki M, Suzuki A. MTCL1 plays an essential role in maintaining Purkinje neuron axon initial segment. *The EMBO journal* 2017; 36: 1227-1242.

E62. Wang B, Zhang M, Ni Y-h, Liu F, Fan H-q, Fei L, Pan X-q, Guo M, Chen R-h, Guo X-r. Identification and characterization of NYGGF4, a novel gene containing a phosphotyrosine-binding (PTB) domain that stimulates 3T3-L1 preadipocytes proliferation. *Gene* 2006; 379: 132-140.

E63. Wu WL, Gan WH, Tong ML, Li XL, Dai JZ, Zhang CM, Guo XR. Over-expression of NYGGF4 (PID1) inhibits glucose transport in skeletal myotubes by blocking the IRS1/PI3K/AKT insulin pathway. *Molecular Genetics and Metabolism* 2011; 102: 374-377.

E64. Zhang C-m, Chen X-h, Wang B, Liu F, Chi X, Tong M-l, Ni Y-h, Chen R-h, Guo X-r. Over-expression of NYGGF4 inhibits glucose transport in 3T3-L1 adipocytes via attenuated phosphorylation of IRS-1 and Akt. *Acta Pharmacologica Sinica* 2008; 30: 120.

E65. Castaldi PJ, Cho MH, Zhou X, Qiu W, McGeachie M, Celli B, Bakke P, Gulsvik A, Lomas DA, Crapo JD, Beaty TH, Rennard S, Harshfield B, Lange C, Singh D, Tal-Singer R, Riley JH, Quackenbush J, Raby BA, Carey VJ, Silverman EK, Hersh CP. Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci. *Human molecular genetics* 2015; 24: 1200-1210.

E66. Dou M, Zhou X, Fan Z, Ding X, Li L, Wang S, Xue W, Wang H, Suo Z, Deng X. Clinical Significance of Retinoic Acid Receptor Beta Promoter Methylation in Prostate Cancer: A Meta-Analysis. *Cellular Physiology and Biochemistry* 2018; 45: 2497-2505.

E67. Richards JB, Waterworth D, O'Rahilly S, Hivert M-F, Loos RJ, Perry JR, Tanaka T, Timpson NJ, Semple RK, Soranzo N. A genome-wide association study reveals variants in ARL15 that influence adiponectin levels. *PLoS genetics* 2009; 5: e1000768.

E68. Paduch M, Jeleñ F, Otlewski J. Structure of small G proteins and their regulators. *ACTA BIOCHIMICA POLONICA-ENGLISH EDITION*- 2001; 48: 829-850.

E69. Rocha N, Payne F, Huang-Doran I, Sleigh A, Fawcett K, Adams C, Stears A, Saudek V, O'Rahilly S, Barroso I. The metabolic syndrome-associated small G protein ARL15 plays a role in adipocyte differentiation and adiponectin secretion. *Scientific reports* 2017; 7: 17593.

E70. Janson C, Kasahara N, Prendergast GC, Colicelli J. RIN3 is a negative regulator of mast cell responses to SCF. *PloS one* 2012; 7: e49615.

E71. Vallet M, Soares DC, Wani S, Sophocleous A, Warner J, Salter DM, Ralston SH, Albagha OM. Targeted sequencing of the Paget's disease associated 14q32 locus identifies several missense coding variants in RIN3 that predispose to Paget's disease of bone. *Human molecular genetics* 2015; 24: 3286-3295.

E72. Silveyra P, DiAngelo SL, Floros J. An 11-nt sequence polymorphism at the 3′ UTR of human SFTPA1 and SFTPA2 gene variants differentially affect gene expression levels and miRNA regulation in cell culture. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2014; 307: L106-L119.

E73. Sorensen GL, Bladbjerg EM, Steffensen R, Tan Q, Madsen J, Drivsholm T, Holmskov U. Association between the surfactant protein D (SFTPD) gene and subclinical carotid artery atherosclerosis. *Atherosclerosis* 2016; 246: 7-12.

E74. Djiadeu P, Farmakovski N, Azzouz D, Kotra LP, Sweezey N, Palaniyar N. Surfactant protein D regulates caspase-8-mediated cascade of the intrinsic pathway of apoptosis while promoting bleb formation. *Molecular immunology* 2017; 92: 190-198.

E75. Schram AW, Baas R, Jansen PW, Riss A, Tora L, Vermeulen M, Timmers HTM. A dual role for SAGA-associated factor 29 (SGF29) in ER stress survival by coordination of both histone H3 acetylation and histone H3 lysine-4 trimethylation. *PloS one* 2013; 8: e70035.

E76. Celedón JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, Reilly JJ, Kwiatkowski DJ, Chapman HA, Laird N, Sylvia JS, Hernandez M, Speizer FE, Weiss ST, Silverman EK. The transforming growth factor-β1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). *Human Molecular Genetics* 2004; 13: 1649-1656.

E77. Poniatowski ŁA, Wojdasiewicz P, Gasik R, Szukiewicz D. Transforming growth factor Beta family: insight into the role of growth factors in regulation of fracture healing biology and potential clinical applications. *Mediators of inflammation* 2015; 2015.

E78. Böttner M, Krieglstein K, Unsicker K. The transforming growth factor-βs: structure, signaling, and roles in nervous system development and functions. *Journal of neurochemistry* 2000; 75: 2227-2240.

E79. Cheng RY, Shang Y, Limjunyawong N, Dao T, Das S, Rabold R, Sham JS, Mitzner W, Tang WY. Alterations of the lung methylome in allergic airway hyper-responsiveness. *Environmental and molecular mutagenesis* 2014; 55: 244-255.

E80. Saito M, Kurokawa M, Oda M, Oshima M, Tsutsui K, Kosaka K, Nakao K, Ogawa M, Manabe R-i, Suda N, Ganjargal G, Hada Y, Noguchi T, Teranaka T, Sekiguchi K, Yoneda T, Tsuji T. ADAMTSL6β Protein Rescues Fibrillin-1 Microfibril Disorder in a Marfan Syndrome Mouse Model through the Promotion of Fibrillin-1 Assembly. *Journal of Biological Chemistry* 2011; 286: 38602-38613.

E81. Tsutsui K, Manabe R-i, Yamada T, Nakano I, Oguri Y, Keene DR, Sengle G, Sakai LY, Sekiguchi K. ADAMTSL-6 Is a Novel Extracellular Matrix Protein That Binds to Fibrillin-1 and Promotes Fibrillin-1 Fibril Formation. *Journal of Biological Chemistry* 2010; 285: 4870-4882.

E82. Panasevich S, Melén E, Hallberg J, Bergström A, Svartengren M, Pershagen G, Nyberg F. Investigation of novel genes for lung function in children and their interaction with tobacco smoke exposure: a preliminary report. *Acta Paediatrica* 2013; 102: 498-503.

E83. Ve T, Vajjhala PR, Hedger A, Croll T, DiMaio F, Horsefield S, Yu X, Lavrencic P, Hassan Z, Morgan GP, Mansell A, Mobli M, O'Carroll A, Chauvin B, Gambin Y, Sierecki E, Landsberg MJ, Stacey KJ, Egelman EH, Kobe B. Structural basis of TIR-domain-assembly formation in MAL- and MyD88-dependent TLR4 signaling. *Nature Structural &Amp; Molecular Biology* 2017; 24: 743.

E84. Bagratuni T, Terpos E, Eleutherakis-Papaiakovou E, Kalapanida D, Gavriatopoulou M, Migkou M, Liacos C-I, Tasidou A, Matsouka C, Mparmparousi D, Dimopoulos MA, Kastritis E. TLR4/TIRAP polymorphisms are associated with progression and survival of patients with symptomatic myeloma. *British Journal of Haematology* 2016; 172: 44-47.

E85. Bernau K, Torr EE, Evans MD, Aoki JK, Ngam CR, Sandbo N. Tensin 1 Is Essential for Myofibroblast Differentiation and Extracellular Matrix Formation. *American journal of respiratory cell and molecular biology* 2017; 56: 465-476.

E86. Haldar D, Agrawal N, Patel S, Kambale PR, Arora K, Sharma A, Tripathi M, Batra A, Kabi BC. Association of VDBP and CYP2R1 gene polymorphisms with vitamin D status in women with polycystic ovarian syndrome: a north Indian study. *European Journal of Nutrition* 2018; 57: 703-711.

E87. Gozdzik A, Zhu J, Wong BYL, Fu L, Cole DEC, Parra EJ. Association of vitamin D binding protein (VDBP) polymorphisms and serum 25(OH)D concentrations in a sample of young Canadian adults of different ancestry. *The Journal of Steroid Biochemistry and Molecular Biology* 2011; 127: 405-412.

E88. Feinstein M, Flusser H, Lerman-Sagie T, Ben-Zeev B, Lev D, Agamy O, Cohen I, Kadir R, Sivan S, Leshinsky-Silver E, Markus B, Birk OS. VPS53 mutations cause progressive cerebello-cerebral atrophy type 2 (PCCA2). *Journal of Medical Genetics* 2014; 51: 303.

E89. Freitag CM, Lempp T, Nguyen TT, Jacob CP, Weissflog L, Romanos M, Renner TJ, Walitza S, Warnke A, Rujescu D, Lesch K-P, Reif A. The role of ASTN2 variants in childhood and

adult ADHD, comorbid disorders and associated personality traits. *Journal of Neural Transmission* 2016; 123: 849-858.

E90. Yin X, Bizon C, Tilson J, Lin Y, Gizer IR, Ehlers CL, Wilhelmsen KC. Genome-wide meta-analysis identifies a novel susceptibility signal at CACNA2D3 for nicotine dependence. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 2017: 10.1002/ajmg.b.32540.

E91. Baeza-Richer C, Arroyo-Pardo E, Blanco-Rojo R, Toxqui L, Remacha A, Vaquero MP, López-Parra AM. Genetic contribution to iron status: SNPs related to iron deficiency anaemia and fine mapping of CACNA2D3 calcium channel subunit. *Blood Cells, Molecules, and Diseases* 2015; 55: 273-280.

E92. Miller S, Melén E, Merid SK, Hall IP, Sayers I. Genes associated with polymorphic variants predicting lung function are differentially expressed during human lung development. *Respiratory Research* 2016; 17: 95.

E93. Grarup N, Andersen G, Krarup NT, Albrechtsen A, Schmitz O, Jørgensen T, Borch-Johnsen K, Hansen T, Pedersen O. Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes. *Diabetes* 2008; 57: 2534-2540.

E94. Berndt A, Leme AS, Shapiro SD. Emerging genetics of COPD. *EMBO molecular medicine* 2012; 4: 1144-1155.

E95. Cioni J-M, Wong HH-W, Bressan D, Kodama L, Harris WA, Holt CE. Axon-Axon Interactions Regulate Topographic Optic Tract Sorting via CYFIP2-Dependent WAVE Complex Function. *Neuron* 2018; 97: 1078-1093.e1076.

E96. Wang Z, Qin G, Zhao TC. HDAC4: mechanism of regulation and biological functions. *Epigenomics* 2014; 6: 139-150.

E97. Barnes PJ. Reduced Histone Deacetylase in COPD: Clinical Implications. *Chest* 2006; 129: 151-155.

E98. Lenormand C, Bausinger H, Gross F, Signorino-Gelo F, Koch S, Peressin M, Fricker D, Cazenave J-P, Bieber T, Hanau D, de la Salle H, Tourne S. HLA-DQA2 and HLA-DQB2 Genes Are Specifically Expressed in Human Langerhans Cells and Encode a New HLA Class II Molecule. *The Journal of Immunology* 2012; 188: 3903.

E99. Siedlinski M, Tingley D, Lipman PJ, Cho MH, Litonjua AA, Sparrow D, Bakke P, Gulsvik A, Lomas DA, Anderson W, Kong X, Rennard SI, Beaty TH, Hokanson JE, Crapo JD, Lange C, Silverman EK, Copdgene, Investigators E. Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Human genetics* 2013; 132: 431-441.

E100. Hancock DB, Wang J-C, Gaddis NC, Levy JL, Saccone NL, Stitzel JA, Goate A, Bierut LJ, Johnson EO. A multiancestry study identifies novel genetic associations with CHRNA5

methylation in human brain and risk of nicotine dependence. *Human molecular genetics* 2015; 24: 5940-5954.

E101. Liu C, Cui H, Gu D, Zhang M, Fang Y, Chen S, Tang M, Zhang B, Chen H. Genetic polymorphisms and lung cancer risk: Evidence from meta-analyses and genome-wide association studies. *Lung Cancer* 2017; 113: 18-29.

E102. Lee H-J, Kim S-Y, Koh J-M, Bok J, Kim K-J, Kim K-S, Park M-H, Shin H-D, Park BL, Kim T-H, Hong JM, Park EK, Kim DJ, Oh B, Kimm K, Kim GS, Lee J-Y. Polymorphisms and haplotypes of integrinα1 (ITGA1) are associated with bone mineral density and fracture risk in postmenopausal Koreans. *Bone* 2007; 41: 979-986.

E103. Abbott GW. The KCNE2 K^+ channel regulatory subunit: Ubiquitous influence, complex pathobiology. *Gene* 2015; 569: 162-172.

E104. Caminos E, Garcia-Pino E, Juiz JM. Loss of auditory activity modifies the location of potassium channel KCNQ5 in auditory brainstem neurons. *Journal of neuroscience research* 2015; 93: 604-614.

E105. Huang H, Trussell LO. KCNQ5 channels control resting properties and release probability of a synapse. *Nature neuroscience* 2011; 14: 840-847.

E106. Georgiev D, González-Burgos G, Kikuchi M, Minabe Y, Lewis DA, Hashimoto T. Selective Expression of KCNS3 Potassium Channel α-Subunit in Parvalbumin-Containing GABA Neurons in the Human Prefrontal Cortex. *PLOS ONE* 2012; 7: e43904.

E107. Georgiev D, Arion D, Enwright JF, Kikuchi M, Minabe Y, Corradi JP, Lewis DA, Hashimoto T. Lower gene expression for KCNS3 potassium channel subunit in parvalbumin-containing neurons in the prefrontal cortex in schizophrenia. *The American journal of psychiatry* 2014; 171: 62-71.

E108. Li J, Peng W, Du L, Yang Q, Wang C, Mo YY. The oncogenic potentials and diagnostic significance of long non-coding RNA LINC00310 in breast cancer. *Journal of Cellular and Molecular Medicine* 2018; 22: 4486-4495.

E109. Tanokashira D, Motoki K, Minegishi S, Hosaka A, Mamada N, Tamaoka A, Okada T, Lakshmana MK, Araki W. LRP1 Downregulates the Alzheimer's β-Secretase BACE1 by Modulating Its Intraneuronal Trafficking. *eNeuro* 2015; 2: ENEURO.0006-0015.2015.

E110. Yang L, Liu C-C, Zheng H, Kanekiyo T, Atagi Y, Jia L, Wang D, N'Songo A, Can D, Xu H, Chen X-F, Bu G. LRP1 modulates the microglial immune response via regulation of JNK and NF-κB signaling pathways. *Journal of neuroinflammation* 2016; 13: 304-304.

E111. Gonias SL, Campana WM. LDL receptor-related protein-1: a regulator of inflammation in atherosclerosis, cancer, and injury to the nervous system. *The American journal of pathology* 2014; 184: 18-27.

E112. Hou A, Zhao L, Zhao F, Wang W, Niu J, Li B, Zhou Z, Zhu D. Expression of MECOM is associated with unfavorable prognosis in glioblastoma multiforme. *OncoTargets and therapy* 2016; 9: 315-320.

E113. Faraco J, Bashir M, Rosenbloom J, Francke U. Characterization of the human gene for microfibril-associated glycoprotein (MFAP2), assignment to chromosome 1p36.1–p35, and linkage to D1S170. *Genomics* 1995; 25: 630-637.

E114. van Wely KH, Molijn AC, Buijs A, Meester-Smoor MA, Aarnoudse AJ, Hellemons A, den Besten P, Grosveld GC, Zwarthoff EC. The MN1 oncoprotein synergizes with coactivators RAC3 and p300 in RAR-RXR-mediated transcription. *Oncogene* 2003; 22: 699.

E115. Pardee TS. Overexpression of MN1 confers resistance to chemotherapy, accelerates leukemia onset, and suppresses p53 and Bim induction. *PLoS One* 2012; 7: e43185.

E116. Mulcahy H, O'rourke K, Adams C, Molloy M, O'gara F. LST1 and NCR3 expression in autoimmune inflammation and in response to IFN-γ, LPS and microbial infection. *Immunogenetics* 2006; 57: 893-903.

E117. Kuek V, Yang Z, Chim SM, Zhu S, Xu H, Chow ST, Tickner J, Rosen V, Erber W, Li X. NPNT is expressed by osteoblasts and mediates angiogenesis via the activation of extracellular signal-regulated kinase. *Scientific reports* 2016; 6: 36210.

E118. Yang L, Liu C-C, Zheng H, Kanekiyo T, Atagi Y, Jia L, Wang D, N'songo A, Can D, Xu H. LRP1 modulates the microglial immune response via regulation of JNK and NF-κB signaling pathways. *Journal of neuroinflammation* 2016; 13: 304.

E119. Toraskar J, Magnussen SN, Chawla K, Svineng G, Steigedal TS. Nephronectin mediates p38 MAPK-induced cell viability via its integrin-binding enhancer motif. *FEBS open bio* 2018; 8: 1992-2001.

E120. Wu J, Yin R-X, Guo T, Lin Q-Z, Shen S-W, Sun J-Q, Shi G-Y, Wu J-Z, Yang D-Z, Lin W-X. Gender-specific association between the cytoplasmic poly (A) binding protein 4 rs4660293 single nucleotide polymorphism and serum lipid levels. *Molecular medicine reports* 2015; 12: 3476-3486.

E121. Fog CK, Galli GG, Lund AH. PRDM proteins: important players in differentiation and disease. *Bioessays* 2012; 34: 50-60.

E122. Loth DW, Artigas MS, Gharib SA, Wain LV, Franceschini N, Koch B, Pottinger TD, Smith AV, Duan Q, Oldmeadow C. Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nature genetics* 2014; 46: 669.

E123. Loth DWaA, Maria Soler and Gharib, Sina A and Wain, Louise V and Franceschini, Nora and Koch, Beate and Pottinger, Tess D and Smith, Albert Vernon and Duan, Qing and Oldmeadow, Chris and Lee, Mi Kyeong and Strachan, David P and James, Alan L and Huffman, Jennifer E and Vitart, Veronique and Ramasamy, Adaikalavan and Wareham, Nicholas J and Kaprio, Jaakko and Wang, Xin-Qun and Trochet, Holly and Kaonen, Mika and Flexeder, Claudia and Albrecht, Eva and Lopez, Lorna M and de Jong, Kim and Thyagarajan, Bharat and Alves, Alexessander Couto and Enroth, Stefan and Omenaas, Ernst and Joshi, Peter K and Fall, Tove and Vinuela, Ana and Launer, Lenore J and Loehr, Laura R and Fornage, Myriam and Li, Guo and Wik, Jemma B and Tang, Wenbo and Manichaikul, Ani and Lahousse, Lies and Harris, Tamara B and North, Kari E and Rudnicka, Alicja R and Hui, Jennie and Gu, Xiangjun and Lumley, Thomas and Wright, Alan F and Hastie, Nicholas D and Campbell, Susan and Kumar, Rajesh and Pin, Isabelle

and Scott, Robert A and Pietilainen, Kirsi H and Surakka, Ida and Liu, Yongmei and Holliday, Elizabeth G and Schulz, Holger and Heinrich, Joachim and Davies, Gail and Vonk, Judith M and Wojczynski, Mary and Pouta, Anneli and Johansson, Asa and Wild, Sarah H and Ingelsson, Erik and Rivadeneira, Fernando and Voezke, Henry and Hysi, Pirro G and Eiriksdottir, Gudny and Morrison, Alanna C and Rotter, Jerome I and Gao, Wei and Postma, Dirkje S and White, Wendy B and Rich, Stephen S and Hofman, Albert and Aspelund, Thor and Couper, David and Smith, Lewis J and Psaty, Bruce M and Lohman, Kurt and Burchard, Esteban G and Uitterlinden, Andre G and Garcia, Melissa and Joubert, Bonnie R and McArdle, Wendy L and Musk, A Bill and Hansel, Nadia and Heckbert, Susan R and Zgaga, Lina and van Meurs, Joyce BJ and Navarro, Pau and Rudan, Igor and Oh, Yeon-Mok and Redline, Susan and Jarvis, Deborah L and Zhao, Jing Hua and Rantanen, Taina and O'Connor, George T and Ripatti, Samuli and Scott, Rodney J and Karrasch, Stefan and Grallert, Harald and Gaddis, Nathan C and Starr, John M and Wijmenga, Cisca and Minster, Ryan L and Lederer, David J and Pekkanen, Juha and Gyllensten, Ulf and Campbe, Harry and Morris, Andrew P and Glaeser, Sven and Hammond, Christopher J and Burkart, Kristin M and Beilby, John and Kritchevsky, Stephen B and Gucinason, Vilrnundur and Hancock, Dana B and Williams, Dale and Polasek, Ozren and Zemunik, Tatijana and Kolcic, Ivana and Petrini, Marcy F and Wjst, Matthias and Kim, Woo Jin and Porteous, David J and Scotland, Generation and Smith, Blair H and Villanen, Anne and Heliovaara, Markku and Attia, John R and Sayers, Ian and Hampel, Regina and Gieger, Christian and Deary, Ian J and Boezen, H Marike and Newman, Anne and Jarvelin, Marjo-Riitta and Wilson, James F and Lind, Lars and Stricker, Bruno H and Teumer, Alexander and Spector, Timothy D and Melen, Erik and Peters, Marjolein J and Lange, Leslie A and Barr, R Graham and Bracke, Ken and Verhamme, Fien and Sung, Joohon and Hiemstra, Pieter S and Cassano, Patricia A and Sood, Akshay and Hayward, Caroline and Dupuis, Josee and Hall, Ian P and Brusselle, Guy and Tobin, Martin D and London, Stephanie J. Genome-wide association analysis identifies six new loci associated with forced vital capacity. *NATURE GENETICS* 2014; 46: 8.

E124. Deng M, Li J, Gan Y, Chen Y, Chen P. Changes in the number of CD31(-)CD45(-)Sca-1(+) cells and Shh signaling pathway involvement in the lungs of mice with emphysema and relevant effects of acute adenovirus infection. *International journal of chronic obstructive pulmonary disease* 2017; 12: 861-872.

E125. Belyaeva OV, Korkina OV, Stetsenko AV, Kedishvili NY. Human retinol dehydrogenase 13 (RDH13) is a mitochondrial short-chain dehydrogenase/reductase with a retinaldehyde reductase activity. *The FEBS journal* 2008; 275: 138-147.

E126. Sakiyama M, Matsuo H, Shimizu S, Nakashima H, Nakayama A, Chiba T, Naito M, Takada T, Suzuki H, Hamajima N, Ichida K, Shimizu T, Shinomiya N. A Common Variant of Organic Anion Transporter 4 (OAT4/SLC22A11) Gene Is Associated with Renal Underexcretion Type Gout. *Drug Metabolism and Pharmacokinetics* 2014; 29: 208-210.

E127. Maran C, Tassone E, Masola V, Onisto M. The Story of SPATA2 (Spermatogenesis-Associated Protein 2): From Sertoli Cells to Pancreatic Beta-Cells. *Current genomics* 2009; 10: 361-363.

E128. Kovanich D, van der Heyden MAG, Aye TT, van Veen TAB, Heck AJR, Scholten A. Sphingosine Kinase Interacting Protein is an A-Kinase Anchoring Protein Specific for Type I cAMP-Dependent Protein Kinase. *ChemBioChem* 2010; 11: 963-971.

E129. Hanafusa T, Mohamed AEA, Domae S, Nakayama E, Ono T. Serological identification of Tektin5 as a cancer/testis antigen and its immunogenicity. *BMC cancer* 2012; 12: 520-520.

E130. Penney KL, Sinnott JA, Tyekucheva S, Gerke T, Shui IM, Kraft P, Sesso HD, Freedman ML, Loda M, Mucci LA, Stampfer MJ. Association of Prostate Cancer Risk Variants with Gene Expression in Normal and Tumor Tissue. *Cancer Epidemiology Biomarkers & Prevention* 2015; 24: 255.

E131. Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, Zhao D, Liu Y, Wang C, Zhang X, Su X, Liu J, Ge W, Levine RL, Li N, Cao X. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* 2015; 525: 389-393.

E132. Yami L, Wen P, Kai Q, Xiaolong L, Zhaolin Z, Jiaojiao C, Dangheng W, Zuo W. TET2: A Novel Epigenetic Regulator and Potential Intervention Target for Atherosclerosis. *DNA and Cell Biology* 2018; 37: 517-523.

E133. Piotr Popławski AUAP-WAUAN. The significance of TRIP11 and T3 signalling pathway in renal cancer progression and survival of patients. *The significance of TRIP11 and T3 signalling pathway in renal cancer progression and survival of patients* 2017; 68: 631-641-631-641.

E134. Subashini C, Dhanesh SB, Chen C-M, Riya PA, Meera V, Divya TS, Kuruvilla R, Buttler K, James J. Wnt5a is a crucial regulator of neurogenesis during cerebellum development. *Scientific Reports* 2017; 7: 42523.

Chapter 7

A randomised-placebo controlled study into the efficacy of working memory training in chronic obstructive pulmonary disease: the study protocol

Authors:

S. Mount, MSc[A] , K. Houben, PhD[B], H. Gosker, PhD[A], M. van Beers MSc[A], L. Schuurman MSc[A], F. Franssen, MD PhD[AC], D. Janssen, MD PhD[CD], A. Schols, PhD[A]

[A] NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Department of Respiratory Medicine, Maastricht, The Netherlands

[B] Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, PO BOX 616, Maastricht 6200 MD, The Netherlands.

[C] Department of Research & Education, Ciro, Horn, the Netherlands.

[D] Centre of Expertise for Palliative Care, Maastricht University Medical Centre+ (MUMC+), Maastricht, the Netherlands.

Abstract

General cognitive impairment is highly prevalent in patients with chronic obstructive pulmonary disease (COPD). Domain-specific cognitive impairments include deficits in working memory (WM), cognitive flexibility, verbal memory, planning and psychomotor speed. These impairments may be associated with poor health behaviours, such as a sedentary lifestyle and low-quality diet. Cognitive training may reverse these effects. Recent evidence suggests that WM training is linked to self-control and, indirectly, to improved lifestyle behaviour including increased physical activity. We aim to investigate the efficacy of WM training (WMT) in patients with COPD on cognitive performance, cognitive stress susceptibility and perception, self-control, and adherence to personalised physical activity and dietary behaviour goals. This trial will be a double blind, placebo-controlled randomised trial will be conducted in 60 patients with COPD. The trial will consist of two phases; 12 weeks of active WM training or sham training followed by 12 weeks of maintenance. Prior to the WMT, before the first phase, participants in both the sham (n=30) and training group (n=30) will set dietary and physical activity goals based on their dietary intake and physical activity profile using validated tools. Cognitive performance will be examined using the Cambridge Neuropsychological Test Automated Battery. The primary outcome of this study will be change in cognitive performance. Secondary outcomes will be self-control (impulsivity), compliance, stress susceptibility and perception, change in dietary intake and daily physical activity level and pattern. This trial will attempt to determine if cognitive performance can be improved in patients with COPD by WMT. Moreover, WM plays a key role in self-regulation of behaviour, i.e. resisting hedonic impulses in exchange for more deliberate evaluations and the achievement of long-term goals. Therefore, we expect that WMT will also have a positive impact on health behaviours.

Registration: Clinicaltrials.gov registration: NCT03073954

Introduction

Chronic obstructive pulmonary disease (COPD) is a serious respiratory condition which affected more that 251 million individuals in 2016,[1] and from 2016 to 2040 is estimated to become the fourth leading cause of life years lost.[2] Chronic obstructive pulmonary disease is characterised by persistent respiratory symptoms and airflow limitation,[3] which is often associated with substantial morbidity and mortality.[4] In addition, COPD patients often suffer from musculoskeletal impairments,[5] cardiovascular comorbidity[6] [7] as well as anxiety [8] and depression.[9] Moreover, COPD patients are at increased risk for cognitive impairment with adverse clinical consequences.[10]

Why are cognitive impairments relevant in COPD?

Impairments in working memory (WM) or other cognitive areas can have a significant impact on patients. Cognitive impairments in general can reduce quality of life, physical activity, social interaction[11] and medication adherence in affected patients.[12] Moreover, it can lead to short-term memory problems, loss of initiative, difficulties with concentration and fatigue.[13] As a result of these consequences, daily activities such as house work may become more difficult,[13] and the risk of all-cause hospitalisation increases.[14] These issues can be further complicated in COPD due to the presence of anxiety and stress. This is particularly relevant to individuals with COPD, as compared to their peers, they experience stronger detrimental mental and quality of life effects when faced with life event related stress.[15] Animal model research suggests that these effects can be further compounded as repeated stress has been shown to cause cognitive impairment [16] and can induce depressive and anxiety like behaviours, in addition to memory deficits.[17] Taken together, this suggests that any existing cognitive impairment may be worsened by inherent stress exposure. Therefore, ways to help temper disease and stress related cognitive impairment are of considerable interest.

Why consider cognitive training in the overall management of COPD?

Cognitive training is an area of translational neuroscience which is gaining interest.[18] The most recent meta-analysis on working memory training (WMT) suggests WTM is effective in mixed populations, but that these effects as well as transfer effects tend to be small.[19] However, these results are primarily from relatively healthy populations, with normal levels of cognitive functioning, and perhaps leave little room for improvement. Individuals with COPD more

frequently experience cognitive impairment;[10] approximately 56.7% of patients with COPD are affected by cognitive impairment, compared to 13.3% of aged matched peers,[10] which can also worsen over time.[20 21] When cognitive function is affected, typically attention, memory, executive functions and WM are affected.[22 23] This is similar to the pattern which has been observed in patients with traumatic brain injury (TBI). In patients with TBI, attention, memory and executive functions are most commonly affected.[24] Interestingly, the mild results observed in mixed populations become more striking if results are focused on patients with a TBI. Specifically, WMT resulted in a moderate and long-lasting improvement in untrained WM tasks, small improvements in evaluations of everyday life functioning, cognitive control and reasoning.[25] Moreover, WMT has been observed to increase prefrontal cortex neural activity and increase the strength of the connectivity between the prefrontal cortex and other brain areas.[26]

Working memory and health behaviours

Recent research has demonstrated the importance of self-control in the regulation of health behaviours, including physical activity.[27] Low levels of self-control are associated with reduced compliance to physical activity[28 29] and healthy diets,[29] obesity, substance abuse, and procrastination.[30] On the other hand, individuals with high self-control are better able to control their thoughts, regulate their emotions and inhibit their impulses.[31]

Self-control is a part of the executive functions, which are a family of top-down mental processes which allow one to concentrate and pay attention to non-instinctual tasks.[32] There are three primary executive functions: inhibition [inhibitory control, self-control (behavioural inhibition) and interference control (selective attention and cognitive inhibition)], WM, and cognitive flexibility.[32 33] Working memory, the ability to select and hold goal-relevant information for a short time, enables us to engage in complex goal-orientated behaviour by managing sensory inputs[18] and plays a key role in cognitive control.[34] Individual differences in WM capacity are related to the ability to inhibit automatic responses, in favour of more opportune controlled-processing responses.[35]

Employing these executive functions costs energy. Simply said it is easier to give in to temptation and to continue to engage in "automatic" reactions than to carefully think about what to do next.[32] The dual process theories of behaviour suggest that the balance between controlled (executive system) and impulsive (impulsive system) behaviour are due to the interaction between these systems.[35] In addition, based on this theory strengthening the executive system could improve goal orientated behaviour by improving control over automatic impulses.[35 36]

The efficacy of specific WMT to improve health behaviours has been demonstrated by a number of studies on eating, smoking, drinking and substance use. Houben et al. observed a reduction in overeating in response to negative emotions and food intake in participants with strong dietary restraint goals,[37] and Dassen et al. showed reduced caloric intake during a sham taste testing trial after a WM intervention, even though no differences in BMI were observed between groups.[38] In smokers, WM capacity was shown to be related to smoking urge and latency to smoke. Specifically, in individuals with poorer WM, urge to smoke was more strongly and negatively associated with smoking latency.[39] Furthermore, WTM has been shown to reduce alcohol intake in problem drinkers [40] as well aid in stabilising street drug and cannabis use in dependent opiate users.[41]

The COGtrain Trial

Lack of physical activity, albeit often not classified as detrimental as illicit drug or alcohol use, can cause considerable damage. Specifically, physical inactivity is a predictor of worse COPD outcomes including progression of exercise intolerance[42] and increased risk of mortality, and is unfortunately frequently observed in patients with COPD.[43] Reducing sedentary behaviour is therefore an important COPD management goal and an integrative part of pulmonary rehabilitation (PR). However, behavioural translation of improved physical capacity after PR to

a more active lifestyle is inconsistent.[44-46] These inconsistencies likely have different drivers but one may be related to low levels of self-control.

Working memory training may enhance self-control by improving attentional control, the efficiency with which attention is regulated towards relevant and away from irrelevant material,[47] and thus aid in maintaining goal relevant information and resisting distraction.[48] These changes could then potentially improve health-related behaviours and in turn could lead to improvements in quality of life. However, it remains to be determined if WMT is effective in patients with COPD, and if it will impact further reaching areas such as behaviour, and if patients will accept online training modules. Therefore, the primary objective of the present clinical trial is to investigate the efficacy of WMT in conjunction with goal setting in patients with COPD on cognitive performance (executive function, episodic memory, visual memory, information processing, and sustained attention). Furthermore, we aim to assess the impact of WMT on self-control (impulsivity), stress susceptibility, perception and compliance to pre-defined individual daily physical activity level and pattern, and dietary advice goals as well as stress.

Hypotheses

1) Working memory training enhances cognitive performance in patients with COPD.
2) Working memory training facilitates the transfer of healthy lifestyle goals to a healthier lifestyle in patients with COPD.
3) Improved cognitive performance reduces stress susceptibility in patients with COPD.

Methods

Patient population

This study will include a population of patients with COPD from the region of South Limburg, the Netherlands. Patients are eligible if they have a diagnosis of COPD based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines[3] and if they are not afflicted by any of the exclusion criteria (Table 1). Once participants have agreed to participate in the study they will be randomly allocated into the treatment or control condition. All measurements will be performed at Maastricht University Medical Centre.

Table 2: Exclusion Criteria

Exclusion criteria
Disease and/or disability limiting ability to undergo neuropsychological testing and/or WMT (e.g. blindness, stroke, or lack of hand control)
Neurological disorders (e.g. Alzheimer's Disease, Parkinson's, Huntington's disease)
Insufficient mastery of the Dutch language
Participation in an inpatient PR programme during study period
Participation in another intervention study during study period

Patients will be recruited using local advertising in newspapers, magazines, and local physiotherapy practices. In addition, patients who have participated in previous studies and have indicated that they may be contacted for future studies, as well as COPD patients visiting the outpatient clinics will be approached.

Intervention

A double blind, placebo-controlled randomised trial will be conducted on 60 patients with COPD. The trial consists of two phases; 12 weeks of intensive training (n=30) or sham training (n=30) (T0-T12) followed by 12 weeks of active follow-up (T12-T24) (figure 1). Of the 30 sessions, patients may miss five sessions. After they have missed three sessions the investigators will be notified and contact the participants. If patients miss more than five training sessions they will be withdrawn from the study. Patients will be evaluated at a screening appointment, baseline, T12 and T24.

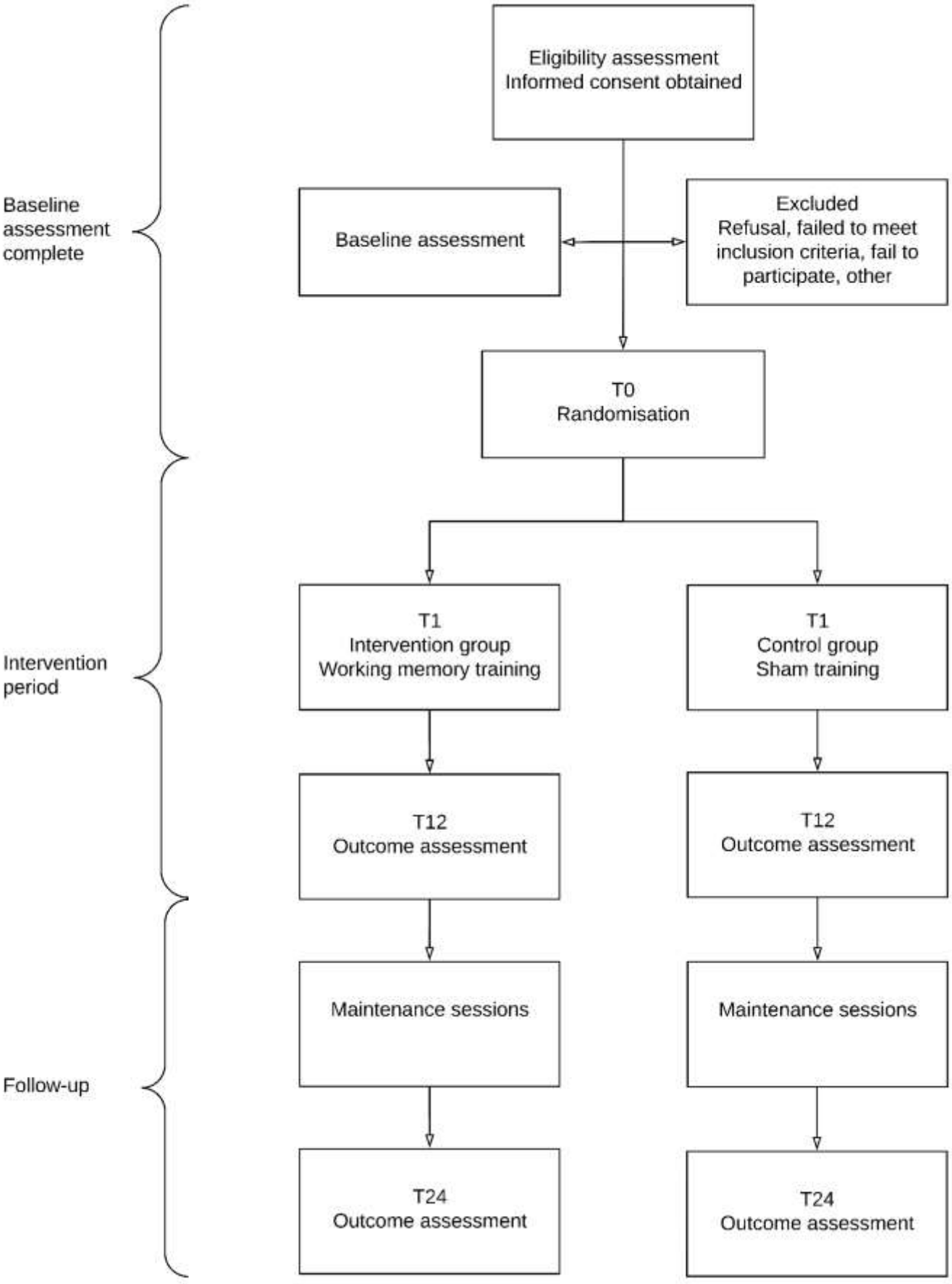


Figure 1: Intervention schedule

Working memory training

The WMT consists of three different tasks: a visuospatial task, a backward digit span task, and a letter span task. In the visuospatial task, participants will be shown a four-by-four grid of squares, some of which flash in blue one after the other. Participants will be required to recall which squares flashed in blue and in which order, by clicking the squares. In the backward digit span task, numbers will be presented on a computer screen, and the participants will be required to reproduce the sequence in the reverse order. Lastly, in the letter span task, letters will be presented one by one in the centre of the screen, and simultaneously with every letter, an accompanying arm will light up (see Figure 2). After all letters have been presented along with their corresponding arms, one arm will light up in red, and participants are required to indicate the letter belonging to that arm on the keyboard of the computer. Participants in the control (sham) group will receive the same tasks as those in the training group but in contrast to the intervention group the tasks will not increase in difficulty by increasing the number of digits, or complexity of the pattern to be recalled.[40]

In the first phase of the study, participants will receive 30 training sessions over a 12-week period, and have to complete at least 25 sessions. Participants will receive a link to every session through e-mail, and have 48 hours to complete a session after receiving the e-mail. In phase two (T12-T24) we will investigate the longer-term effects of the intervention. There is evidence to support the maintenance of the training effect after the cessation of the intervention.[50-53] However, providing booster sessions could greatly enhance the long-term effects of the training. Ball et al.[54] demonstrated that one booster session compensated for nearly five months of cognitive decline, and the positive results of the training intervention

were apparent for years after it ended. Given the potential benefits of booster sessions we will therefore offer the participants one booster session per week for three months in the maintenance phase (phase two) after completion of the training.

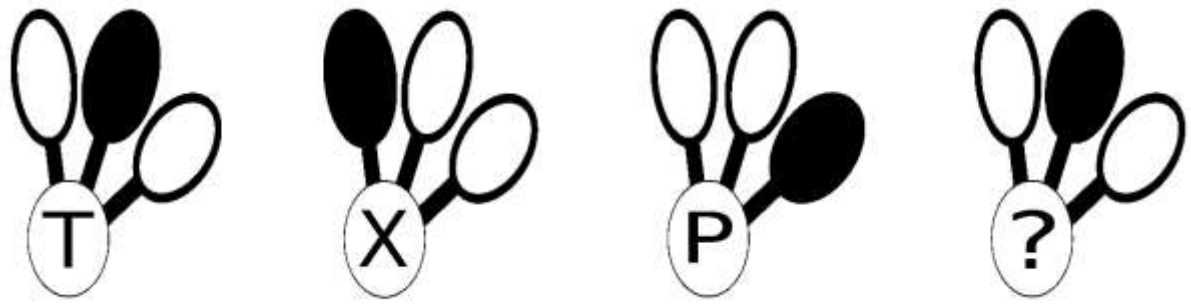


Figure 2: Example of working memory training exercise

Goal setting

Behaviour with respect to physical activity and dietary intake will be measured at T0 and at the end of phase 1 (T12) and 2 (T24) in both the intervention and the control group. Physical activity data will be collected by an ActivPAL™ accelerometer. This device provides a well-established measurement of both physical activity and sedentary time. Subjects wear the accelerometer fastened to their leg with Tegaderm™ adhesive tape for 7 consecutive days. The ActivPAL™ calculates body posture as sitting/lying, standing, and stepping and energy expenditure (METs) using static and dynamic acceleration information.[55] Energy expenditure can then be classified from the ActivPAL™ into sedentary (G1.5 METs), light (1.5–2.99 METs), and MVPA (>3 METs) intensity.[55]

Dietary intake will be monitored using a 24-hour recall questionnaire in the form of an interview. Participants will be asked if the past 24 hours were reflective of typical dietary intake. If dietary intake varied significantly from normal, participants will be asked to recall their intake on the previous day.

After having analysed the accelerometery data and the dietary questionnaire, participants will be informed about the results by a trained research assistant. Dietary and/or physical activity goals will then be set by the patients together with the research assistant. Representative scores will be calculated based on the Alternative Healthy Eating Index (AHEI)[56] to aid participants in understanding how they can improve their diets. Information will be presented as a score and in graphical form for individual categories such as fruit, vegetable, red meat, excessive alcohol consumption, among others, for easy interpretation and guided goal setting discussions. Physical activity will be presented as daily step counts, and percent time spent in sitting, standing and moving activities in the form of graphs. The graphs will also clearly show the time of day which type of activity was performed. Goals can include dietary changes such as reducing alcohol, red meat, increasing whole grains, fruit and or vegetable consumption; physical activity will be in the form of steps per day.

Study parameters and endpoints

The study parameters are listed in Table 2. Study participants will visit twice, separated by a week, before the study (T0) to determine baseline performance. Additionally, they will be

tested at the end of phase one after 12 weeks (T12) and at the end of phase two after 24 weeks (T24). All measurements will be taken by the trained investigators.

Primary outcome: measures of cognitive function

The Cambridge Neuropsychological Test Automated Battery (CANTAB) is a widely used cognitive function assessment tool which has been used in a large range of clinical and non-clinical studies. Using the CANTAB system, we will test WM, cognitive flexibility, and planning with the motor screening task, paired associated learning task, reaction time task, delated matching to sample task and a spatial WM task. Furthermore, we will administer the Stop-Signal Task (SST) as a measure of impulsivity.

As an additional measure of cognitive function, we administer the Addenbrooke's Cognition Examination Revised (ACE-R); a brief test battery which assesses five domains, namely: orientation and attention, memory, verbal fluency, language and visuospatial ability. The ACE-R is a comprehensive screening tool, and has good psychometric properties: both sensitivity and specificity are around 0.9.[57-59]

Secondary measurement outcomes

Additional outcome measures will include a measure of perceived stress (Cohen's Perceived Stress Scale (PSS)),[60] chronic stress (hair cortisol),[61] acute stress (salivary cortisol awakening response (CAR))[62] stress response (socially evaluated cold pressor test (SECPT)),[63] functional

exercise capacity (6 minute walk test),[64] physical performance (short performance test battery (SPBB)),[65] disease-specific health status (COPD assessment test, (CAT)),[66] motivation for exercise and dietary intake (Behavioral Regulation in Exercise Questionnaire-2 (BREQ-2)),[67] the Regulation of Eating Behavior Scale REBS,[68] depression (Beck depression inventory(BDI-II)),[69] anxiety (generalised anxiety disorder – 7 (GAD-7))[70] and dietary intake (Food frequency questionnaire (FFQ)).

Table 2

Primary outcomes	*Instrument*	*T0*	*T1*	*T12*	*T24*
Cognitive performance					
	Cambridge Neuropsychological Test Automated Battery	X*	X	X	X
	The Addenbrooke's Cognitive Examination Revised		X		
Secondary outcomes					
Cognitive stress susceptibility and perception					
Chronic stress	The Perceived Stress Scale		X	X	X
	Hair cortisol		X	X	
	Salivary cortisol awakening response		X	X	
Acute stress	Socially Evaluated Cold pressor test		X	X	X
Physical activity					
	Accelerometer: step count, gait variability		X	X	X
	6-minute walking test	X	X	X	X
Balance					
	Short Performance Battery		X	X	X
Quality of life					
	COPD assessment test		X	X	X
Motivational questionnaires					
	Behavioural Regulation in Exercise Questionnaire-2	X			X
	Regulation of Eating Behaviour Scale	X			X
Psychological wellbeing					
	Beck Depression Inventory-second edition		X	X	X
	Generalised Anxiety Disorder-7		X	X	X
Dietary intake					
	Food Frequency Questionnaire	X		X	X

Table 2 continued

Other characteristics

Socioeconomic variables					
	Age, gender, education level	X[$]	X[$]		
Anthropometry					
	Height	X[$]	X[$]		
	Bioelectrical impedance	X[$]	X[$]	X	X
	Waist circumference	X[$]	X[$]		
	Weight	X[$]	X[$]	X	X
Other clinical characteristics					
	Smoking status, exacerbations, COPD Gold classification, spirometry	X			
Medication					
	Questionnaire		X	X	X
Manipulation Check					
	Manipulation check			X	X
Compliance and accessibility					
	Training compliance			X	X

Measurement timing of study parameters
* Required to compensate for any possible learning effects.
[$] Measurements may be taken at screening or T1.

Manipulation check

This will be done in the form of a very short structured interview in which the participant will be asked to recall and name the specific dietary and physical activity goals they made at the beginning of the intervention. This test will be administered in a way similar to that of Hatchell and colleagues,[71] where participants are asked to recall key health messages from their personalised healthy lifestyle advice sessions. Patient responses will be recorded in writing. The recalled points will be compared to the personalised advice given to the participant. Responses will then be scored as follows: 0 points – field blank or no recall of the message content; 1 point – key points not directly related to the message themes; 2 points – key points

directly related to the message themes.[71] Patients who recall more information will be given higher scores.

Compliance and accessibility

Training compliance will serve as a measure of compliance and accessibility of the WMT. During the trial, patient participation in the online sessions will be recorded. Here we can examine participant engagement to the training (total time spent), number of completed sessions, answer patterns as well as monitor attrition rate.

Statistical Analysis

Data analyses will be conducted using the statistical package IBM SPSS Statistics for Windows, version 21.0 (SPSS, Inc., Chicago, IL). All subjects who complete the study will be included in the analysis of the primary outcome. Missing data will be considered as missing at random and will not be imputed. Two-sided *p*-values less than 0.05 will be claimed as statistically significant. No interim analysis will be performed.

Demographic and clinical background information including socioeconomic status, age, gender, education level, smoking status, alcohol intake, medication use, oxygen therapy, exacerbations, comorbidities, COPD Gold classification, and spirometry will be collected at baseline.

To determine the effects of WMT on our primary and secondary parameters, assuming the data meets the requirements, a Repeated Measures Analysis of Variance (baseline compared to Post-intervention and 3-month secondary follow-up) will be used to compare mean changes. If the data are non-normally distributed, and transformation is deemed inappropriate, a generalised linear mixed model approach will be used instead. Statistical measures of the interaction between time points and group will be reported. Effect sizes will be reported as Cohen's d, computed as the difference in performance at baseline and post-intervention or 6-month follow-up between the two groups. Effect sizes of 0.8 are considered large and effect sizes between 0.5 and 0.8 are moderate. If participants withdraw from the study, their data will be used up to the point of their withdrawal, although techniques such as multiple imputation will not be used to deal with missing data.

Sample size and power

The sample size calculation was calculated with G*Power 3.1.9.2 and sample sizes used in the most comparable studies from literature. In a recent Canadian study examining the effects of cognitive training on cognitive decline, the authors anticipated the effect size of the training would be f = 0.475.[72] When taking into account an α of 0.05 and a power of 95% these parameters result in a required sample size of 60 individuals, or 30 per group. To maintain an appropriate sample size for analysis efficacy, new patients will be recruited to compensate for patients who drop out of the study until 60 participants have completed the study.

Randomisation

The randomisation will be performed by an independent researcher via www.randomization.com, before the participants start the training. A randomisation block will be used with 70 subjects randomised in 7 blocks (10 per block, 5x control and 5x active training) (allowing for drop out). After 40 subjects have been randomised, the same independent researcher will verify if the distribution between the groups on the basis of age and gender is similar. All researchers involved will remain blinded until the completion of the study and analysis.

Data management and monitoring

Participant data is stored on a secured network server accessible only to the researchers. All paper documents are stored in a secured cabinet located at Maastricht University Medical Centre. Data is entered into the database by the researchers or research assistants, which is then periodically examined by the data monitor. After completion of the study the database will be cleaned and compared to original documentation in the case of obscure values.

Data monitoring is performed by the independent Clinical Trials Centre Maastricht (CTCM) committee at trial commencement, trial closing and twice during the course of the study.

Discussion

We hypothesise that it is possible to improve WM in patients with COPD. Moreover, because WM plays a key role in this self-regulation of behaviour, i.e. resisting hedonic impulses in

exchange for more deliberate evaluations and the achievement of long-term goals, we expect that WMT will also have a positive influence on diet and physical activity.

Our study has a number of strengths and weaknesses. A first strength is that we use a double-blind placebo-controlled design. Second, we monitor patient compliance carefully using the online training programme and compliance (completing at least 25 out of 30 training sessions) is a prerequisite for continuing the trial. Third, by approaching patients with COPD in the community, physiotherapy and outpatient clinics, we simultaneously test the efficacy and the feasibility of implementing online training modules in a broad COPD patient population. Fourth, we actively measure transfer of WM training. In addition, we have a long patient follow-up period which allows us to determine the longer-term effects of WM training, as well as compliance to the offered booster session. Lastly, in addition to examining the possible cognitive benefits of WM training, and providing healthy lifestyle tips and goals we are examining the broader impact of the online training, specifically if it aids in adopting a healthier lifestyle. On the other hand, our study is limited by the patient recruitment area, which may limit the external validity of our findings to other international populations. Secondly, only motivated patients are likely to participate in this study. Moreover, we have some risk of inclusion bias due to the requirement of regular access to a computer and an internet connection. In addition, there may be a relatively high risk of drop out in this study due to the repetitive nature of the tasks and time required.

Anticipated clinical implications

If the results of this study are positive it would indicate that WMT should be adopted in COPD management given the prevalence of (mild) cognitive impairment in COPD patients. Adding WMT would be a simple addition to any programme as patients can complete the training without supervision either in a home or in a rehabilitation setting given. Future updates will likely allow this type of training to be performed on tablets or smart mobile phones and offline. Positive results in this study also imply a greater need to create awareness among patients with COPD and their caregivers and physicians to remain cognitively active, for instance through reading newspapers or playing cognitively challenging games such as chess or sudoku.

Trial status

The trial started on 19 October 2017. At the time of submission, the recruitment was ongoing and will presumably be completed in June 2020. This study follows the SPIRIT guidelines.

Abbreviations

AHEI	Alternative Healthy Eating Index
ACE-R	Addenbrooke's Cognition Examination Revised (ACE-R)
BDI-II	Beck depression inventory
BREQ-2	Behavioural Regulation in Exercise Questionnaire-2
CANTAB	Cambridge Neuropsychological Test Automated Battery
CAR	Cortisol awakening response
CAT	COPD assessment test
COPD	Chronic obstructive pulmonary disease
CTCM	Clinical Trials Centre Maastricht
FFQ	Food frequency questionnaire
GAD-7	Generalized anxiety disorder – 7
GOLD	Global Initiative for Chronic Obstructive Lung Disease
METC	Ethics Committee of Maastricht University
METS	Metabolic equivalents
MUMC+	Maastricht University Medical Centre+
MVPA	Moderate to vigorous activity bouts
N	Number
PR	Pulmonary rehabilitation
PSS	Cohen's Perceived Stress Scale
REBS	Regulation of Eating Behavior Scale
SECPT	Socially evaluated cold pressor test
SPBB	Short performance test battery
SST	Stop signal task

TBI	Traumatic brain injury
T0-T24	Time 0, Time 1 week, Time 12 weeks, Time 24 weeks
WM	working memory
WMT	working memory training

Declarations

Acknowledgements

The researchers would like to thank the future and current participants of this study as well as the independent physician Prof. dr. Anne-Marie Dingemans.

Funding Statement

This work was supported by the Maastricht University Interfaculty Programme 'Eatwell', and through the JPI HDHL project "Ambrosiac" by ZonMW (project number 529051006).

Author's Contributions

SM did the literature search and did the statistical power calculation in consultation with AS. SM wrote the first draft of the article. KH, MvB, FF, DJ, HG and AS gave valuable input in drafting the manuscript. In collaboration with MvB, SM wrote the medical ethical protocol for the study, which was also carefully reviewed by KH, FF, DJ, HG and AM. All authors critically revised the manuscript for intellectual content, finally approved of the version to be published, and agree to be held accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethical considerations

The study was approved by the local Ethics Committee of Maastricht University (METC) in June 2017 (NL59883.068.17/ METC173010). Any required amendments to the study will be submitted for approval by the local accredited METC, and implemented after favourable opinion by the METC. No measurements will be carried out before written and informed consent has been obtained by the investigators. Participants can withdraw from the study at

any point. The total burden of participation consists of approximately 30 online cognitive training sessions of 20 to 30 minutes each, in the form of a game on a mobile device or computer, as well as outcome assessment at baseline and after 12 weeks. In addition, participants will be asked to complete maintenance online cognitive training sessions, once per week during the 3-month follow-up period.

Dissemination

The findings of the study will be disseminated through peer-reviewed journals, national and international conference presentations and to the COPD patients through a newsletter and/ or presentation.

Patient and Public Involvement statement

This study was designed to meet an unmet need, to improve adherence to patient rehabilitation programmes and thus patient outcomes. Patients were not involved in the direct design of the study but patient burden was considered carefully by the researchers. Furthermore, patients are not involved in recruitment and conduct of the study beyond their own participation and personal responses to study advertisements. Study results will be communicated with participants as discussed in the dissemination section.

Competing interest's statement

Authors declare no conflicting interests.

References

1. Organization WH. Chronic obstructive pulmonary disease (COPD) Fact sheet Geneva: World Health Organization; 2017 [December 2017:[Available from: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) accessed 28/01 2019.
2. Foreman KJ, Marquez N, Dolgert A, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet* 2018;392(10159):2052-90.
3. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management and prevention of chronic obstructive lung disease 2017 report. *Respirology* 2017;22(3):575-601.
4. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet* 2013;380(9859):2197-223.
5. Cielen N, Maes K, Gayan-Ramirez G. Musculoskeletal disorders in chronic obstructive pulmonary disease. *BioMed research international* 2014;2014
6. Cavaillès A, Brinchault-Rabin G, Dixmier A, et al. Comorbidities of COPD. *European Respiratory Review* 2013;22(130):454-75.
7. Chen W, Thomas J, Sadatsafavi M, et al. Risk of cardiovascular comorbidity in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *The Lancet Respiratory medicine* 2015;3(8):631-39.
8. Willgoss TG, Yohannes AM. Anxiety disorders in patients with COPD: a systematic review. *Respiratory care* 2013;58(5):858-66.
9. Schneider C, Jick SS, Bothner U, et al. COPD and the risk of depression. *Chest* 2010;137(2):341-47.
10. Cleutjens FA, Franssen FM, Spruit MA, et al. Domain-specific cognitive impairment in patients with COPD and control subjects. *International journal of chronic obstructive pulmonary disease* 2017;12:1.
11. Yohannes AM, Chen W, Moga AM, et al. Cognitive impairment in chronic obstructive pulmonary disease and chronic heart failure: a systematic review and meta-analysis of observational studies. *Journal of the American Medical Directors Association* 2017;18(5):451. e1-51. e11.
12. Campbell NL, Boustani MA, Skopelja EN, et al. Medication adherence in older adults with cognitive impairment: a systematic evidence-based review. *The American journal of geriatric pharmacotherapy* 2012;10(3):165-77.
13. Johansson MM, Marcusson J, Wressle E. Cognitive impairment and its consequences in everyday life: experiences of people with mild cognitive impairment or mild dementia and their relatives. *International psychogeriatrics* 2015;27(6):949-58.
14. Chang SS, Chen S, McAvay GJ, et al. Effect of coexisting chronic obstructive pulmonary disease and cognitive impairment on health outcomes in older adults. *Journal of the American Geriatrics Society* 2012;60(10):1839-46.
15. Lu Y, Nyunt MSZ, Gwee X, et al. Life event stress and chronic obstructive pulmonary disease (COPD): associations with mental well-being and quality of life in a population-based study. *BMJ open* 2012;2(6):e001674.

16. Yuen EY, Wei J, Liu W, et al. Repeated stress causes cognitive impairment by suppressing glutamate receptor expression and function in prefrontal cortex. *Neuron* 2012;73(5):962-77.
17. Gawali NB, Bulani VD, Gursahani MS, et al. Agmatine attenuates chronic unpredictable mild stress-induced anxiety, depression-like behaviours and cognitive impairment by modulating nitrergic signalling pathway. *Brain research* 2017;1663:66-77.
18. Simon AJ, Skinner SN, Ziegler DA. Training Working Memory: Anatomy Matters. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2016;36(30):7805-06. doi: 10.1523/JNEUROSCI.1513-16.2016
19. Soveri A, Antfolk J, Karlsson L, et al. Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review* 2017;24(4):1077-96. doi: 10.3758/s13423-016-1217-0
20. Hung WW, Wisnivesky JP, Siu AL, et al. Cognitive decline among patients with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2009;180(2):134-37.
21. Park SK. Trajectories of change in cognitive function in people with COPD. *Journal of clinical nursing* 2018
22. Dodd J, Getov S, Jones PW. Cognitive function in COPD. *European Respiratory Journal* 2010;35(4):913-22.
23. Mora L, Sagot C, Dieudonne B, et al. Link between non hypoxemic chronic obstructive pulmonary disease and executive functioning in oldest old. *Geriatrie et psychologie neuropsychiatrie du vieillissement* 2015;13(3):335-42.
24. Arciniegas DB, Held K, Wagner P. Cognitive impairment following traumatic brain injury. *Current Treatment Options in Neurology* 2002;4(1):43-57.
25. Weicker J, Villringer A, Thöne-Otto A. Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology* 2016;30(2):190.
26. Constantinidis C, Klingberg T. The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience* 2016;17(7):438.
27. Hagger MS, Wood C, Stiff C, et al. The strength model of self-regulation failure and health-related behaviour. *Health Psychology Review* 2009;3(2):208-38.
28. Martin Ginis KA, Bray SR. Application of the limited strength model of self-regulation to understanding exercise effort, planning and adherence. *Psychology and Health* 2010;25(10):1147-60.
29. Wills TA, Isasi CR, Mendoza D, et al. Self-control constructs related to measures of dietary intake and physical activity in adolescents. *Journal of Adolescent Health* 2007;41(6):551-58.
30. de Ridder DT, Lensvelt-Mulders G, Finkenauer C, et al. Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review* 2012;16(1):76-99.
31. Tangney JP, Baumeister RF, Boone AL. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of personality* 2004;72(2):271-324.
32. Diamond A. Executive functions. *Annual review of psychology* 2013;64:135-68.
33. Miyake A, Friedman NP, Emerson MJ, et al. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology* 2000;41(1):49-100.

34. Hofmann W, Schmeichel BJ, Baddeley AD. Executive functions and self-regulation. *Trends in Cognitive Sciences* 2012;16(3):174-80. doi: https://doi.org/10.1016/j.tics.2012.01.006
35. Dassen FC, Houben K, Van Breukelen GJ, et al. Gamified working memory training in overweight individuals reduces food intake but not body weight. *Appetite* 2018;124:89-98.
36. Hofmann W, Friese M, Strack F. Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science* 2009;4(2):162-76.
37. Houben K, Dassen FC, Jansen A. Taking control: Working memory training in overweight individuals increases self-regulation of food intake. *Appetite* 2016;105:567-74.
38. Dassen FC, Houben K, Van Breukelen GJ, et al. Gamified working memory training in overweight individuals reduces food intake but not body weight. *Appetite* 2017
39. Day AM, Kahler CW, Metrik J, et al. Working memory moderates the association between smoking urge and smoking lapse behavior after alcohol administration in a laboratory analogue task. *Nicotine & Tobacco Research* 2014;17(9):1173-77.
40. Houben K, Wiers RW, Jansen A. Getting a grip on drinking behavior: training working memory to reduce alcohol abuse. *Psychological science* 2011;22(7):968-75.
41. Rass O, Schacht RL, Buckheit K, et al. A randomized controlled trial of the effects of working memory training in methadone maintenance patients. *Drug and alcohol dependence* 2015;156:38-46.
42. Troosters T, Gosselink R, Janssens W, et al. Exercise training and pulmonary rehabilitation: new insights and remaining challenges. *European Respiratory Review* 2010;19(115):24-29.
43. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. *American journal of respiratory and critical care medicine* 2013;188(8):e13-e64.
44. Cindy Ng LW, Mackney J, Jenkins S, et al. Does exercise training change physical activity in people with COPD? A systematic review and meta-analysis. *Chronic respiratory disease* 2012;9(1):17-26.
45. Troosters T, Demeyer H, Hornikx M, et al. Pulmonary Rehabilitation. *Clinics in Chest Medicine* 2014;35(1):241-49. doi: https://doi.org/10.1016/j.ccm.2013.10.006
46. Egan C, Deering BM, Blake C, et al. Short term and long term effects of pulmonary rehabilitation on physical activity in COPD. *Respiratory medicine* 2012;106(12):1671-79.
47. Duncan J, Humphreys GW. Visual search and stimulus similarity. *Psychological review* 1989;96(3):433.
48. Unsworth N, Spillers GJ. Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language* 2010;62(4):392-406.
49. Putra AC, Tanimoto K, Arifin M, et al. Genetic variations in detoxification enzymes and HIF-1α in Japanese patients with COPD. *The clinical respiratory journal* 2013;7(1):7-15.
50. Vermeij A, Claassen JA, Dautzenberg PL, et al. Transfer and maintenance effects of online working-memory training in normal ageing and mild cognitive impairment. *Neuropsychological rehabilitation* 2016;26(5-6):783-809.
51. Brehmer Y, Westerberg H, Bäckman L. Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in human neuroscience* 2012;6

52. Carretti B, Borella E, Zavagnin M, et al. Gains in language comprehension relating to working memory training in healthy older adults. *International journal of geriatric psychiatry* 2013;28(5):539-46.
53. Borella E, Carretti B, Zanoni G, et al. Working memory training in old age: an examination of transfer and maintenance effects. *Archives of clinical neuropsychology* 2013;28(4):331-47.
54. Ball KK, Ross LA, Roth DL, et al. Speed of processing training in the ACTIVE study: how much is needed and who benefits? *Journal of aging and health* 2013;25(8_suppl):65S-84S.
55. Lyden K, Keadle SK, Staudenmayer J, et al. The activPALTM Accurately Classifies Activity Intensity Categories in Healthy Adults. *Medicine and science in sports and exercise* 2017;49(5):1022-28.
56. Chiuve SE, Fung TT, Rimm EB, et al. Alternative dietary indices both strongly predict risk of chronic disease. *The Journal of nutrition* 2012:jn. 111.157222.
57. Lonie JA, Tierney KM, Ebmeier KP. Screening for mild cognitive impairment: a systematic review. *International journal of geriatric psychiatry* 2009;24(9):902-15.
58. Fang R, Wang G, Huang Y, et al. Validation of the Chinese version of Addenbrooke's cognitive examination-revised for screening mild Alzheimer's disease and mild cognitive impairment. *Dementia and geriatric cognitive disorders* 2014;37(3-4):223-31.
59. Pigliautile M, Ricci M, Mioshi E, et al. Validation study of the Italian Addenbrooke's Cognitive Examination Revised in a young-old and old-old population. *Dementia and Geriatric Cognitive Disorders* 2011;32(5):301-07.
60. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of health and social behavior* 1983:385-96.
61. O'Brien KM, Tronick E, Moore CL. Relationship between hair cortisol and perceived chronic stress in a diverse sample. *Stress and Health* 2013;29(4):337-44.
62. Wust S, Wolf J, Hellhammer DH, et al. The cortisol awakening response-normal values and confounds. *Noise and health* 2000;2(7):79.
63. Schwabe L, Haddad L, Schachinger H. HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology* 2008;33(6):890-95.
64. Holland AE, Spruit MA, Troosters T, et al. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease: Eur Respiratory Soc, 2014.
65. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *Journal of gerontology* 1994;49(2):M85-M94.
66. Jones P, Harding G, Berry P, et al. Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009;34(3):648-54.
67. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. *Journal of Sport and Exercise Psychology* 2004;26(2):191-96.
68. Pelletier LG, Dion SC, Slovinec-D'Angelo M, et al. Why do you regulate what you eat? Relationships between forms of regulation, eating behaviors, sustained dietary behavior change, and psychological adjustment. *Motivation and Emotion* 2004;28(3):245-77.
69. Beck AT, Steer RA, Ball R, et al. Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of personality assessment* 1996;67(3):588-97.

70. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 2006;166(10):1092-97.
71. Hatchell AC, Bassett-Gunter RL, Clarke M, et al. Messages for men: The efficacy of EPPM-based messages targeting men's physical activity. *Health Psychology* 2013;32(1):24.
72. Gill DP, Gregory MA, Zou G, et al. The healthy mind, healthy mobility trial: a novel exercise program for older adults. *Med Sci Sports Exerc* 2016;48(2):297-306.

Chapter 8
General Discussion

Personalised medicine is the concept that we can or will be able to customise treatments for each and every person in order to achieve better health outcomes. For example, a patient may have a genetic variation that increases the half-life of medications, leading to adverse effects, while another patient might have a genetic variation that makes the drugs less effective. This indicates that the best treatment option for these patients is different, even if they have a similar clinical presentation.[1] This personalisation trend is something we are beginning to see in complex diseases like chronic obstructive pulmonary disease (COPD). Although COPD treatments are not yet down to the level of a person's genetic blueprint, they are beginning to be customised to the clinical phenotypes: the non-exacerbator with either chronic bronchitis or emphysema, the asthma–COPD overlap syndrome, the frequent exacerbator with emphysema, or the frequent exacerbator with chronic bronchitis.[2] These phenotypes however, may not only be the result of specific genetic predispositions, they may be in fact the result of gene-environment interactions. That is to say, the effect of a genotype on disease risk is different in persons with different environmental exposures. Regardless of the specific cause, if it's purely genetic or if it's gene-environment interactions, the development of so-called precision/personalised medicine for COPD is impeded due to the lack of understanding of the specific disease aetiology.

Genome wide association studies (GWAS) may help us to understand the aetiology of chronic diseases, such as COPD, and could generate new leads for drug discoveries. The first GWAS was published a short 14 years ago in 2005,[3] [4] following the mapping of the genome in April 2003. From that point onwards, over a thousand GWAS have been published on many different

diseases, including COPD. However, despite the success of the GWAS in identifying genetic variations associated with COPD, these variations only explain a fraction of the heritability of this disease.[3] The so called *"missing heritability"* might be caused by gene-environment interactions. Using COPD we can clearly demonstrate this principle by discussing the risk associated with the interaction between cigarette smoke and the C-1562T variation (rs3918242) in the MMP-9 gene.[5] Specifically, functional analyses of this gene showed significantly higher MMP-9 promoter activity in the T promoter compared to the C promoter after exposure to cigarette smoke condensate.[5] Matrix metalloproteinases (MMP's) are endopeptidases that catalyse the degradation of extra cellular matrix components and activate growth factors, receptors and adhesion molecules.[6] Although the exact contribution of MMP-9 to the development of COPD is not well understood, cigarette smoking had a significant effect on lung function of T-allele carriers.[5] This suggests increased macrophage MMP-9 activity related tissue degradation as a result of cigarette smoke exposure. Unfortunately, such functional analyses are not yet available of all potential genetic risk variations for COPD. However, the search for the specific aetiologies of COPD continues and new functional analyses are expected in the near future. These updates, hopefully, include functional analyses of genetic risk variations under the influence of other environmental exposures, such as environmental pollutants and poor diet since these factors are largely modifiable. Specifically, although genetics is likely to play an important role in COPD aetiology, our meta- analysis (Chapter five) showed that as much as 87 percent of COPD cases could be avoided by adapting an optimal lifestyle. Therefore, future research should focus on both genetic as well as lifestyle and genetic-lifestyle interaction, in order to completely understand the aetiology of COPD and thereby, finding new leads to prevent and treat this disease. Undoubtedly, discoveries here will also be applicable to other lifestyle related chronic diseases.

Although, GWAS are very powerful hypothesis free tools (thereby overcoming the obstacles imposed by the incomplete understanding of disease pathophysiology) to identify variants underlying common disorders, we also need to recognise the criticisms they face. An important concern is the translation of novel genetic associations into medical innovations. The identification of novel genetic variations is strongly dependent on the study size and is not necessarily a causal variant for the disease in question.[7] In order to identify the causal variants, knowledge of every single genetic variation associated with the disease of interest is required. Therefore, meta-meta-analyses on identified genetic variations, as presented in Chapter five of this thesis, as well as network analysis to put the genetic variations into their biological context, as presented in Chapter six, are essential. In our meta-meta-analyses, we ordered the COPD associated genetic variations based on their relevance, i.e. the strength of the associated risk, making it possible to better identify risky genetic variants. This will allow researchers to better develop prediction models as well as targeted personalised interventions and or preventative treatments. In chapter six we not only showed the likely genetic culprits with respect to the development of COPD, but also how minor changes in other perhaps less relevant variations could affect risk on a systems level. Additionally, by using variant effect prediction (VEP) analysis we looked even more closely at the genetic variations associated with COPD risk, and saw that only a handful of the identified genetic variations associated with COPD actually resulted in functional changes, and thus convey actual changes in risk. Again, this shows that the results of meta-analyses should be interpreted with care as they can produce a number of statistically significant false leads. By combining the results of larger meta-analyses with tools like network and VEP analysis laboratory studies and prediction models can be better designed.

As mentioned previously, and showed in Chapter five, COPD is mainly caused by lifestyle factors, therefore, knowing which specific lifestyle risk factors, beyond just smoking, will allow us to better predict who and who will not development COPD. This in-turn, allows for better healthcare personalisation. However, due to the lack of evidence for lifestyle risk factors influencing COPD development, current models for predicting COPD perform poorly.[8] The current best performing model only incorporates risk factors such as age, gender, low birthweight, early life respiratory infection, and in addition, includes the presence of specific risk genetic variations.[8] Although this model performs reasonably well, the results of our meta-meta-analysis and network analysis suggests that the prediction model could be significantly improved by adding the genetic variations causing a functional change, and by considering environmental exposures such as western diet, smoking behaviour, exposure to solid fuel smoke and poor ventilation. Until we can predict with 100 per cent certainty, or entirely prevent exposure to environmental and or behavioural risk factors, we need to continue to develop better models in order to allow medical doctors to impact human health.

Predicting how and when someone may develop COPD is just part of the successful ager puzzle, i.e. who will get sick, when and how it can be prevented; we also need to be able to effectively treat COPD patients. COPD is an accelerated ageing syndrome affecting not only the lungs but also extrapulmonary tissues.[9] [10] Although incurable, COPD can be well managed if patients follow an extensive treatment regime, which often involves medication, dietary changes and exercise. With the latter two comes a significant amount of difficulty for the patient. While using medication may be easy, changes in dietary habits and physical activity level may be difficult, particularly for those who have been inactive throughout their lives and follow a lifelong poor diet. Research has shown that only a minor fraction of the COPD patients

manages to sustain behavioural changes in diet and exercise upon their release from pulmonary rehabilitation.[11-13] The lack of behaviour change may not be due to intentional deviance but due to lack of self-control. Therefore, when it comes to novel treatments methods to aid patients in improving compliance to treatment regimens, cognitive barriers are important to consider and tackle.

In Chapter seven we propose "working memory training (WMT)" as a novel intervention. This training may aid patients sticking to healthier lifestyle choices, as the working memory (WM) is part of the executive functions top-down control of mental processes, which are required to avoid automatic, instinctual, or intuitional responses and allow you to concentrate and pay attention to tasks or important stimuli.[14-16] Recent research has established that WM contributes to self-regulation of behaviour, including eating behaviour.[17] Self-regulation is the ability to actively inhibit behavioural responses and impulses that are incompatible with one's goals.[17] In this process, the WM serves as an active mental representation of self-regulatory goals, by top-down control of attention away from stimuli and toward goal-relevant information, by suppressing ruminative thoughts, down-regulating of unwanted affect, desires and cravings and shielding goals and standards from interference.[17] However it remains controversial if WMC training can result in near or far transfer effects, i.e. if training one cognitive area can result in improvements in other non-trained cognitive functions.[18-20] Specifically it is questioned if WMC training can result in near and or far transfer effects on behavioural inhibition; an active inhibition of habitual behaviours and prepotent impulses.[17]

Self-control is a limited resource, which can become depleted through engaging in tasks or activities which make high demands on self-control.[21 22] As we discuss in Chapter seven, there is evidence that working memory can be expanded. Given patients with COPD often suffer from some degree of cognitive impairment[23] with the executively functions particularly being affected,[24 25] that COPD is largely a lifestyle disease, and that sustained improvement in daily physical activity following PR is uncommon,[11-13] this trial should it be successful, may serve as a new adjunctive treatment avenue for COPD and perhaps other chronic disease patients.

COPD is just one of many chronic life-style related diseases which can be considered as syndromes of failing to age successfully, and putting one at an increased risk of becoming frail. Since the number of aged individuals has exponentially increased over the last decade, as does the number of aged individuals living with one or more chronic disease, successful ageing has become one of the main areas of interest in public health. PubMed lists in excess of 6500 articles related to successful ageing, with over 350 articles being published this year (i.e. 2019), making it extraordinarily difficult keeping up in this specific research field. We therefore, reviewed all current scientific literature on this topic performed between 2013 and 2016 in order to gain insight into modern advances in this field (Chapter two). Results of this review shows that in order to determine what it is to age successfully, as well as to define the concept of successful ageing remains a serious challenge for both health care professionals and researchers. Different models, including different outcomes are proposed for the concept of successful ageing. Although no consensus has been made, a clear change in model focus (i.e. from mortality and frailty to maintained physical and cognitive abilities) has been identified over the years. Moreover, developments have been restricted due to the technological

limitations of the available cohorts, that the available data was unsuitable to detect early and subtle declines in function and/or were inappropriate measures for younger old adults.

The healthy ageing phenotype (HAP) was proposed in 2013 as an answer, or partial answer to the lack of appropriate successful ageing outcome markers. This model, developed through the work of a consortium and further refined and published by Lara *et a,l*[26] suggests that successful ageing can be operationalised and defined as having high levels of psychological wellbeing, social wellbeing, physiological and metabolic health, physical capability and cognitive function.[26] To determine the validity of this healthy ageing phenotype, a confirmatory factor analysis was performed on a large dataset from a Dutch cohort study; *"the Maastricht Study (TMS)"* (Chapter three). Results show that this HAP could not be validated within this dataset due to methodological issues (data was non-factorable) which led us to try a data driven method using both The Maastricht Study dataset (TMS) as well as the Italian *"InChianti* "data set (Chapter four). This data driven approach in the InChianti dataset showed a four-factor model for ageing successfully, including neuro-sensory function, muscle function, cardiovascular function and an adiposity domain. Moreover, it showed that these domains contributed to walking speed as well as the ability to perform activities of daily living at baseline and after the nine-year follow-up period, both of which are key indicators of maintained function i.e. successful ageing. Not only do these results give interesting insight with respect to the interrelationships between variables, they also highlight the importance of sensory function and maintained physical ability (walking speed, balance, chair stands etc.).

In Chapter four, we further tested the TMS data set and found a two-factor successful ageing model, factor one (HbA1c, HDL, pulse pressure, BMI, MMSE, GIT, WLTR, processing speed, executive function, emotional support, contact frequency, TCST time, WT speed) and factor two (GAD, Aggression, personal mastery, GSES persistence, GSES initiative). Differences between the results of the TMS and the InChianti study may be the result of differences in variables included in the model. For example, the InChianti data set contained variables describing sensory function were available while the TMS did not. Differences could also have been the result of methodological differences. In the InChianti study, we pre-selected a number of variables as outcome parameters prior to performing the factor analysis procedure due to the longitudinal availability of the data. In the TMS, we selected variables purely based on those suggested by the healthy ageing phenotype (HAP) and did not try to make any such predictions. Thus, resulted in variables such as those describing measures of emotional vitality to be available for factor analysis in the TMS study but not for the InChianti study. Given these methodological differences it is unsurprising that the results differed significantly. Moreover, there were strong differences in the populations used for these studies. The InChianti study is an Italian cohort and ranges in age from 20-102 whereas the TMS population is Dutch, aged 40-75 and had a large proportion of diabetic individuals. Recent research also suggests that different populations may age differently. More specifically, a recent study which examined population level age-related morbidity and mortality by aggregating all disability-adjusted life years (DALYs; a measurement of loss of healthy life, related to the 92 diseases), across 195 countries between 1990 and 2017,[27] showed that the Dutch population ranked considerably more poorly (27th in 2017) than Italy (fifth 2017).[27] This suggests a difference in age-related disease burden between the countries, which may also explain differences in the factor structures.

The differences we observed between these two study populations highlights the need for personalised medicine and healthcare. Here we observed two clearly different factor structures which may not only have been influenced by the design of the cohort study but also from the inherent differences between these populations, cultural ideals, healthcare availability and behavioural factors. We need a clear objective definition of healthy ageing which have be operationalised to measure ageing health such that it is useable across cultures. This would allow for the detection of early warning signs, as well as for research to be more comparable across countries. This in turn would aid in the assessment of personalised interventions across cultures.

Ethical considerations

With the development of big data also considerable other challenges and dilemmas needs to be addressed. One such challenge is *"how to deal with the ethical dilemmas that come along with big data and personalized healthcare"*. With respect to data, researchers need to be able to answer the questions of what, who, where, when and how. What data will be collected, by whom? Who will have access to it? For how long? Where and until when will it be stored? When can we or health professionals access this information and how will they do this? How will privacy be guaranteed? How will discrimination and abuse of this data be prevented? These questions may seem trivial at first but the answers become considerably more complicated the longer they are considered due to the complexity ensuring optimised care while at the same time guarding against data misuse and discrimination. Answering these questions, however, is key to the development of big data. Each and everyone has a genetic blueprint which guides

our development. This is however just part of the story as gene-environment interactions play a strong role in how each person's genetic story develops.

In an ideal world, with respect to the development of personalised therapies, all information could be anonymously and simultaneously collected and matched with biomarkers, as well as health outcomes, without the risk of discrimination. However even in this idealistic world, ethical questions would remain. Such as if or how people should be informed about specific risks, and how to cope with individuals who do not wish to know this information or how to cope with the rights family members? As big data gets bigger, we need to carefully consider these implications.

Other challenges that comes along with big data are the high-dimensionality of biomedical data, incomplete, biased, heterogeneous, dynamic, and noisy.[28] Although, research has already suggested ways to deal with these problems, it will remain a challenge as the complexity of data increases. Hopefully, one day in our near future, we will be able to predict successful ageing with a 100 per cent certainty.

References

1. Brookfield JFY. Q&A: promise and pitfalls of genome-wide association studies. *BMC biology* 2010;8:41-41.
2. Sidhaye VK, Nishida K, Martinez FJ. Precision medicine in COPD: where are we and where do we need to go? *European Respiratory Review* 2018;27(149):180022.
3. Ikegawa S. A short history of the genome-wide association study: where we were and where we are going. *Genomics & informatics* 2012;10(4):220.
4. Thomas DC, Haile RW, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. *The American Journal of Human Genetics* 2005;77(3):337-45.
5. Stankovic M, Kojic S, Djordjevic V, et al. Gene-environment interaction between the MMP9 C–1562T promoter variant and cigarette smoke in the pathogenesis of chronic obstructive pulmonary disease. *Environmental and molecular mutagenesis* 2016;57(6):447-54.
6. Greenlee KJ, Werb Z, Kheradmand F. Matrix metalloproteinases in lung: multiple, multifarious, and multifaceted. *Physiological reviews* 2007;87(1):69-98.
7. Visscher PM, Brown MA, McCarthy MI, et al. Five years of GWAS discovery. *The American Journal of Human Genetics* 2012;90(1):7-24.
8. Matheson MC, Bowatte G, Perret JL, et al. Prediction models for the development of COPD: a systematic review. *International Journal of Chronic Obstructive Pulmonary Disease* 2018;13:1927.
9. MacNee W. Is chronic obstructive pulmonary disease an accelerated aging disease? *Annals of the American Thoracic Society* 2016;13(Supplement 5):S429-S37.
10. Karametos I, Tsiboli P, Togousidis I, et al. Chronic Obstructive Pulmonary Disease as a Main Factor of Premature Aging. *International journal of environmental research and public health* 2019;16(4):540.
11. Egan C, Deering BM, Blake C, et al. Short term and long term effects of pulmonary rehabilitation on physical activity in COPD. *Respiratory medicine* 2012;106(12):1671-79.
12. Cindy Ng LW, Mackney J, Jenkins S, et al. Does exercise training change physical activity in people with COPD? A systematic review and meta-analysis. *Chronic respiratory disease* 2012;9(1):17-26.
13. Troosters T, Demeyer H, Hornikx M, et al. Pulmonary rehabilitation. *Clinics in chest medicine* 2014;35(1):241-49.
14. Burgess PW, Simons JS. 18 Theories of frontal lobe executive function: clinical applications. *The effectiveness of rehabilitation for cognitive deficits* 2005:211.
15. Espy KA, Sheffield TD, Wiebe SA, et al. Executive control and dimensions of problem behaviors in preschool children. *Journal of Child Psychology and Psychiatry* 2011;52(1):33-46.
16. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* 2001;24(1):167-202.
17. Hofmann W, Schmeichel BJ, Baddeley AD. Executive functions and self-regulation. *Trends in Cognitive Sciences* 2012;16(3):174-80. doi: https://doi.org/10.1016/j.tics.2012.01.006
18. Melby-Lervåg M, Redick TS, Hulme C. Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer": Evidence From a Meta-Analytic Review. *Perspect Psychol Sci* 2016;11(4):512-34. doi: 10.1177/1745691616635612
19. Spencer-Smith M, Klingberg T. Benefits of a working memory training program for inattention in daily life: a systematic review and meta-analysis. *PloS one* 2015;10(3):e0119522-e22. doi: 10.1371/journal.pone.0119522
20. Soveri A, Antfolk J, Karlsson L, et al. Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review* 2017;24(4):1077-96.
21. Alves CRR, Tessaro VH, Teixeira LAC, et al. Influence of Acute High-Intensity Aerobic Interval Exercise Bout on Selective Attention and Short-Term Memory Tasks. *Perceptual and Motor Skills* 2014;118(1):63-72. doi: 10.2466/22.06.PMS.118k10w4

22. Dishman RK. Motivating older adults to exercise. *South Med J* 1994;87(5):S79-82.
23. Cleutjens FA, Franssen FM, Spruit MA, et al. Domain-specific cognitive impairment in patients with COPD and control subjects. *International journal of chronic obstructive pulmonary disease* 2017;12:1.
24. Dodd J, Getov S, Jones PW. Cognitive function in COPD. *European Respiratory Journal* 2010;35(4):913-22.
25. Mora L, Sagot C, Dieudonne B, et al. Link between non hypoxemic chronic obstructive pulmonary disease and executive functioning in oldest old. *Geriatrie et psychologie neuropsychiatrie du vieillissement* 2015;13(3):335-42.
26. Lara J, Godfrey A, Evans E, et al. Towards measurement of the Healthy Ageing Phenotype in lifestyle-based intervention studies. *Maturitas* 2013;76(2):189-99. doi: http://dx.doi.org/10.1016/j.maturitas.2013.07.007
27. Chang AY, Skirbekk VF, Tyrovolas S, et al. Measuring population ageing: an analysis of the Global Burden of Disease Study 2017. *The Lancet Public Health* 2019;4(3):e159-e67. doi: 10.1016/S2468-2667(19)30019-2
28. Zitnik M, Nguyen F, Wang B, et al. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion* 2019;50:71-91. doi: https://doi.org/10.1016/j.inffus.2018.09.012
29. Mendes-Soares H, Raveh-Sadka T, Azulay S, et al. Assessment of a Personalized Approach to Predicting Postprandial Glycemic Responses to Food Among Individuals Without DiabetesPersonalized Approach to Predicting Postprandial Glycemic Responses to FoodPersonalized Approach to Predicting Postprandial Glycemic Responses to Food. *JAMA Network Open* 2019;2(2):e188102-e02. doi: 10.1001/jamanetworkopen.2018.8102

Conclusion

In this thesis, several approaches, including factor analysis (Chapter three & four), meta-analysis (Chapter five) and network analysis (Chapter six), have been taken to integrate findings from different research areas (i.e. genetics, lifestyle, gene-environment interaction, and behaviour), in order to get one step closer to understand the complexity of human health and ageing. However, since almost every day new and exciting findings are generated in different fields of research, more powerful analytic tools need to be developed and validated to deal with this so-called big data and that effectively integrate findings from different research areas.

Data mining and pattern recognition are such methods that have the potential to unravel important patterns in complex data. In both methods computers uses statistical models and algorithms without any specific instructions. Understanding these complex inter-relationships will not only help us define healthy ageing, but also to examine both modifiable as well as non-modifiable determinants of ageing, as well as their interactions (i.e. gene-environment) such that we can customise health care regimes and optimise preventative heath assessments. However, with such developments the end user needs to be kept in mind. A metric needs to be comprehensive and should ideally be able to be used in combination with both genetic and behavioural data so that it may be possible to determine who will get sick when, with what, and when they are likely to develop frailty risk factors, while maintaining some of the simplicity inherent to simple risk scores such as BMI. This will allow individuals as well as clinicians to understand how risk is developed but also how it can be mitigated.

Although reductionist models such as BMI are useful due to their inherent simplicity, it is also this characteristic which makes them obsolete, with respect to disease risk prediction. One only has to consider the obesity paradox, or the fact that not all diseases are caused by the most relevant exposures. A key example of this is smoking, it does not always lead to the development of COPD (or other chronic diseases including cancer). This emphasises the importance of cumulative risks; chronic disease is usually the result of multiple factors, including genetics, environment and behaviour. Yes, individual risk factors, unquestionably increase disease risks, however considering such limited models, with perhaps the addition of gender, education and age is simply not enough when we are facing catastrophic changes in our population pyramids.

Simplifications of risk scores are key to aid in public understanding as well as to avoid confusion. Moreover, they may help physicians communicate more effectively to their patients given time constraints. On the other hand, over simplification can ignore vital information and ignore new developments. In the case of BMI, we know it can be deceiving as it is often used as a static measurement, and does not give an indication of adiposity or muscle mass. Similarly, in the case of glycaemic index, another simplified model which indicates how blood sugar levels will respond to specific foods, recent research has shown that more accurate glycaemic responses can be produced when previous meals, age, gender, microbiome among other characteristics are accounted for.[29] Thus we need models which cannot only effectively communicate specific health messages, they also need to be user friendly for the patient and the physician.

In the end though, it remains to be questioned if this ever-increasing demand for complexity, bigger, better models, really matters? Is a good diet, healthy body weight, sufficient exercise within healthy social circles not enough to help us avoid most diseases and age successfully? When they are not, when we fail to meet our physical and psychological needs and overstep the boundaries as to what health requires, i.e. to be human and less than perfect and less than perfect situations, we will have ever improving prediction models to determine who will get sick, when and with what. This hopefully will help us to prevent common chronic diseases and aid in helping us to age successfully – a state of maintained cognitive, physical, psychological and physiological wellbeing.

Summary

The relevance of ageing successfully is becoming increasingly important as life expectancy increases and birth rates fall. Successful ageing means maintaining health and wellbeing into old age. Meaning, by aiding individuals to age well, we can prevent millions of premature deaths, chronic disease, disease related disability and years with reduced quality of life. Therefore, the ability to measure such factors influencing the life course is an important public health concern. Risk factors for disease and dependency can be considered non-controllable, distal, and intermediate. By developing an effective metric considering these factors, we could not only predict who is likely to age successfully, but also measure the effectiveness of interventions, and improve and personalise interventions in order to maximise the therapeutic effects.

Measures of successful ageing

There is great interest in developing tools to measure healthy ageing as well as in identifying the early stages of health impairment. However, there is lack of consensus as to how to exactly define healthy ageing. Unsurprisingly, due to the lack of clear definition, there is also significant diversity in so called healthy ageing assessment tools. Despite controversy, progress is being made in describing and devising tools to capture the healthy ageing phenotype. Attempts to measure healthy ageing have relied primarily on cross-sectional data collected in older people. More recent studies assessed the healthy ageing phenotype by using markers of multiple functional domains and used longitudinal data to model the dynamics and trajectories of healthy ageing. These attempts, however, relied on markers and data from earlier cohort studies and are limited by the tools used to collect data in those studies. Such data are often

unsuitable to detect early subtle declines in function and/or are inappropriate for use in younger old adult populations.

Measuring successful ageing

The development of tools to measure ageing has been limited by the lack of appropriate outcome measures, and operational definitions of successful ageing. In this thesis, however, we proposed to measure successful ageing, or at least a proxy of it by designing a tool that includes representative variables of physical function, cognitive status, social interactions, psychological status, blood biomarkers, disease history, and socioeconomic status. For this, we used data driven methods and found a four-domain health model, including; neuro-sensory function, muscle function, cardio-metabolic function and adiposity. This model could predict walking speed and dependency at baseline and longitudinally over a nine-year period. Unfortunately, the same model was not able to predict self-rated health or emotional vitality, thereby, suggesting a multi-domain health model can be used to predict objective but not subjective measures of successful ageing.

Other models of healthy ageing

There are many models of healthy ageing which have been proposed, but few have been tested. We tested whether a multidimensional model based on systematic review of literature and expert opinion, the Healthy ageing phenotype (HAP), was an appropriate operationalization of healthy ageing in a Dutch population. We used cross-sectional data from the Maastricht Study (TMS) and selected variables based on the HAP five domain model (i.e. cognitive function, social wellbeing, physical capability, psychological wellbeing, and physiological and metabolic health). Among individuals from the south of the Netherlands,

aged between 40 and 70, we discovered that, although this model makes sense theoretically, data could not be combined in this way using statistical analysis, indicating that this model does not fit the data. Subsequent exploratory analysis suggested a two-domain model, including physiological, cognitive, social, physical capability domain, as well as a psychological domain may have been more appropriate in this population.

COPD and Accelerated ageing

COPD is considered a model of accelerated ageing as it exemplifies the key features of ageing including telomere shortening, cellular senescence, activation of PI3 kinase-mTOR signalling, impaired autophagy, mitochondrial dysfunction, stem cell exhaustion, epigenetic changes, abnormal microRNA profiles, immune senescence, and low-grade systemic inflammation. Moreover, the risk factors for COPD are similar to many other chronic diseases. Smoking is the most important risk factor for COPD. Not all smokers, however, develop COPD, suggesting that other lifestyle-, environmental- and genetic factors may play a role in the development of COPD. In this thesis, we reviewed all available meta-analyses reporting on genetic variants, lifestyle or environmental factors associated with development of COPD. For genetic variants, we found 42 relevant publications, including data on 281 genetic variants. Of these 281, 74% (n=208) showed to be significantly associated with COPD, with COPD with odds ratio's ranging from 0.17 to 3.33. For lifestyle/environmental factors, 11 relevant publications were identified and reported on exposures to various types of pollution such as exposure to vapours, cigarette smoke etc. as well as poor diet. Of the 281 genetic variants we identified, 74 percent (n=208), and 87 percent (seven) of the lifestyle or environmental factors showed a significant association with COPD with odds ratio's ranging from 0.17 to 3.33 and 0.45 to 9.50 respectively.

Network analysis

The list of gene mutations associated with disease risk has little value outside of prediction models unless they are considered within biological systems. By examining genetic factors in a biological context, individual genetic risks or single nucleotide polymorphisms (SNP's), as well as possible therapeutic targets, can be better predicted. In this thesis, we were able to look at the SNP's associated with COPD more closely, by making use of network analysis and variant effect prediction analysis. As a first step, we classified genes into 11 different functional classes such as detoxification and cellular metabolism based on their (hypothesised) function. This allowed us to more easily examine the SNP's and their associated genes, and study which genes were most likely to have significant environmental interactions. Of the 315 identified biological pathways, derived from 181 statistically significant genetic risk factors for COPD, we found that only seven had a potentially influential mutation, namely in *AK9, SERPINA1, IL27, CYP1A1, EPHX1, SLC22A11 and TESMIN*. Functional analysis of genes highlighted that of all the identified mutations, only mutations in the genes involved in inflammatory and detoxification pathways are likely to be relevant with respect to COPD risk. Furthermore, these analysis show that more emphasis needs to be placed on gene-environment, gene-behaviour and gene-lifestyle interactions, instead of looking only at predispositions.

Interventions in unsuccessful agers

Although a great deal of work needs to be done, with respect to better disease prediction models, effective interventions are still highly needed for those who have already aged unsuccessfully and developed chronic disease. For COPD, key parts of these interventions involve lifestyle changes, improving diet, exercise and smoking cessation. However, historically changes adapted in therapeutic settings are poorly adhered to once the active intervention

has ended. Exercise and healthy diet are key for prevention and maintenance of good health; this is relevant for COPD patients, but is also important for maintaining health and preventing other chronic illnesses. However, when inactivity and poor diet behaviours are deeply ingrained, change is difficult to achieve and or maintain. Recent evidence suggests that working memory training may aid in self-regulation and the adherence to healthy lifestyle goals, thereby reducing sedentary activity and improving dietary habits. This is supported by recent evidence in other problem populations in which working memory training was shown to reduce alcohol intake in problem drinkers. We hypothesise that this type of cognitive training will also be beneficial in COPD patients, and that it may also help them to reduce sedentary activity and improve dietary intake.

Valorisation

Successful ageing

If the twentieth century was the century of population growth, then the twenty-first century will become the century of ageing.[1] The number of old and very old adults (aged 65 and over, and 80 and over respectively) is rapidly rising in all European countries, and represents a progressively growing percentage of the general population.[2] At the same time, the proportion of working aged individuals is declining.[3] These changes in the population pyramid, as well as increasing life expectancy, is challenging the stability of health and social care systems.[4] Furthermore, age-related diseases are mounting as a result of healthy life expectancy lagging behind increased life expectancy.[1]

Biological ageing varies markedly between individuals,[5] and this disparity between individuals only grows with age.[6] Although partially genetically determined, 75 percent of human longevity is believed to be determined by modifiable factors including diet, lifestyle and socioeconomic status.[5] In order to understand whether any intervention aimed at promoting healthy ageing is effective, a benchmark for the assessment of healthy ageing is needed. Due to the multidimensional nature of ageing, and age-related pathologies, assessing healthy ageing by combining information across many different measurements would seem to be the solution.

In the past and maybe the present, health has been over simplified into single units of measure. Given this over simplification there is substantial room for the development of personalised health. This is particularly true in an era when we have the ability to measure everything that

matters, from activity trackers, sleep, vital signs, blood pressure, heart rate and stress and develop algorithms in combination with genetic and physiologic information for the purposes of making personalised recommendations.[7] We need to consider the synergisms and interactions between different aspects of human life in a broader sense. By understanding the dynamic ways in which we age, the ability to differentiate between exceptional disease-free ageing from one associated with increased frailty and decreased quality of life, will not only aid us in the comprehending disease and ageing processes but also grants us the opportunity to develop targeted therapies. However significant progress in this field has been hampered by a lack of consensus on the definition of successful and or healthy ageing.[8]

Societal and economic relevance

Ageing and disease are intertwined but they do not have to be. But studying ageing gives us an insight into disease, and disease insight into ageing. Recent developments in scientific literature suggest it may be possible and realistic to slow the ageing process.[9] Delaying ageing could increase life expectancy by 2.2 years most of which could be spent in relatively good health, while saving 7.1 trillion dollars over the next fifty years according to an American simulation study.[9] The same efforts put into heart disease and cancer treatment would result in declining returns as improvements in health and longevity would diminish by 2060 according the same model.[9] Moreover, if chronic metabolic diseases were dealt with using appropriate dietary strategies, statistical models estimate the death insurance claims would drop by 13 percent, meaning a reduction in premature loss of life from preventable conditions.[10] Therefore, efforts to establish a definition, metric and benchmark of optimal or healthy ageing could potentially save trillions of dollars through the use of early targeted preventative health interventions.

Relevance of new research methodology

Our research shows that complex measurements can be combined into meaningful metrics using data driven methods, and that these metrics are predictive of the risk of developing frailty. Specifically, we show that walking speed and dependency risk can be predicted over a nine-year period using a combined measurement. Moreover, we emphasise the usefulness of using data driven methods in this field from our failure to validate the Healthy Ageing Phenotype model that was based on systematic review of available literature. Not only does this method allow us to make use of a wide array of data, it demonstrates the potential value of future research using machine learning techniques.

The sequencing of the human genome opened the door to characterising traits of health and disease and linking it to genetic information.[11] However the function of many genes remains unknown, and what is known, is limited to a few cell types, tissues or physiological contexts.[11] Difficulties in information collection, differences in disease manifestations, descriptive phenotypes or disease subclasses requires extensive examination of the discrete components of disease phenotypes, information which is not typically recorded in medical charts and further complicates elucidation.[11] Delving into this information may help to link seemingly unrelated conditions which share common biological pathways and/or disease mechanisms[11] or help us better understand the process of cellular and organismal senescence.

By carefully examining the genes most likely to be involved in disease processes, such as with meta-analyses it helps to weed out some of the statistical noise; the genes shown to be statistically associated with the development of a particular gene, but in reality, only show statistical significance due to the large number of genes tested. When this shortened list of genes then further examined within a biological context, where we can see how gene and gene products potentially interact in a system, we can make better hypotheses about which genes are most and least likely to be involved in the disease process. Furthermore, by studying how individual suspect genes, i.e. the ones with strong risk associations and plausible biological relationships, have changed from their 'natural' form, we can examine what the downstream implications may be. From these, focused traditional laboratory studies can be designed, the results of which can help in the development of new treatments and therapies.

Combining measurements, predictive models, trackers to genetics, understanding how these interactions work, and what the consequences are, helps us to develop targeted preventative treatments and therapies. Moreover, it will help us to understand how, when and in whom disease will develop. Assuming a positive ethical environment, one where such information will not be abused, and only used for preventive care, potentially saving thousands of dollars in healthcare costs in addition to improving quality of life. However, even if all treatments in the end are optimised, key is in prevention, and once a disease has developed, prevention of progression is also of the utmost importance. Until structural changes of disease can be reversed, such as alveolar wall destruction in emphysema, most therapies in chronic obstructive pulmonary disease are merely symptom control. Therefore, available and new strategies to maintain or improve lifestyle remain relevant particularly given the difficulty in

adhering to healthy lifestyles. In patients with COPD for example, even after pulmonary rehabilitation (PR), where emphasis is put on exercise training and physical activity, the uptake of a more active lifestyle upon the completion of PR is inconsistent.[12-14] However, it is unlikely that these drivers are unique to chronic diseases such as COPD. Impediments to healthier lifestyles are likely almost universal in nature and may include low levels of self-control. Any research improving adherence will greatly improve health status by reducing the risks of inactivity, namely progression of exercise intolerance[15] and increased risk of mortality.[16] We present a trial to improve self-control through working memory training. Should this trial be successful in improving working memory in COPD patients, and if this translates into improved self-control, physicians may be able to add another tool to regular PR therapy. Moreover, it represents a cost effect treatment option for other lifestyle related chronic diseases.

Translation into practice

Combining new definitions, metrics and interventions will bring us further into the future. By defining health, we can measure it. By measuring it we can predict who, what, and where a disease or risk of disease will develop. With predictions we can intervene to mitigate further risks and hopefully improve quality of life while reducing health costs, and by improving interventions, we can help people adhere to healthier lifestyles. Taken together this research can be further explored in different diseases and therapeutic areas to improve prediction models and potentially disease therapies, should our trial prove to be effective. Which can act together to improve quality of life but also aid in reducing health care costs, lost productive time, and curb the affects population ageing.

References

1. Greaves CJ, Sheppard KE, Abraham C, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. *BMC public health* 2011;11(1):119.
2. Commission E. People in the EU – statistics on an ageing society Luxembourg: Eurostat; 2017 [updated 23 October 2017. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/People_in_the_EU_–_statistics_on_an_ageing_society accessed 23 October 2017.
3. Commission E. Population structure and ageing Luxembourg: Eurostat; 2017 [updated 20 July 2017. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing accessed 23 October 2017.
4. Berrío Valencia MI. Aging population: a challenge for public health. *Revista Colombiana de Anestesiología* 2012;40(3):192-94.
5. Collerton J, Barrass K, Bond J, et al. The Newcastle 85+ study: biological, clinical and psychosocial factors associated with healthy ageing: study protocol. *BMC geriatrics* 2007;7(1):14.
6. Martin P, Kelly N, Kahana B, et al. Defining successful aging: A tangible or elusive concept? *The Gerontologist* 2014;55(1):14-25.
7. Ordovas JM, Ferguson LR, Tai ES, et al. Personalised nutrition and health. *BMJ* 2018;361:bmj. k2173.
8. Michel J-P, Graf C, Ecarnot F. Individual healthy aging indices, measurements and scores. *Aging clinical and experimental research* 2019:1-7.
9. Goldman DP, Cutler D, Rowe JW, et al. Substantial health and economic returns from delayed aging may warrant a new focus for medical research. *Health affairs* 2013;32(10):1698-705.
10. Schoonbee J, Elefteriadis E. Why would a leading global reinsurer be interested in nutrition? *BMJ* 2018;361:k2435.
11. Delude CM. The details of disease. *Nature* 2015;527(7576):S14.
12. Egan C, Deering BM, Blake C, et al. Short term and long term effects of pulmonary rehabilitation on physical activity in COPD. *Respiratory medicine* 2012;106(12):1671-79.
13. Cindy Ng LW, Mackney J, Jenkins S, et al. Does exercise training change physical activity in people with COPD? A systematic review and meta-analysis. *Chronic respiratory disease* 2012;9(1):17-26.
14. Spruit MA, Pitta F, McAuley E, et al. Pulmonary rehabilitation and physical activity in patients with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2015;192(8):924-33.
15. Troosters T, Gosselink R, Janssens W, et al. Exercise training and pulmonary rehabilitation: new insights and remaining challenges. *European Respiratory Review* 2010;19(115):24-29.
16. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. *American journal of respiratory and critical care medicine* 2013;188(8):e13-e64.

List of publications

Field, DT, Bell, L., **Mount, S,** Williams, CM, & Butler, LT. (2014). Flavonoids and visual function: Observations and hypotheses. In Handbook of Nutrition, Diet and the Eye (pp. 403-411). Academic Press.

Mount, S, Lara J, Schols AMWJ, Mathers JC. Towards a multidimensional healthy ageing phenotype. Current Opinion in Clinical Nutrition and Metabolic Care 19.6 (2016): 418-426.

Mount, S, Ferrucci, L., Wesselius, A, Zeegers, MP, & Schols, AMWJ. (2019). Measuring successful aging: an exploratory factor analysis of the InCHIANTI Study into different health domains. Aging.

Mount, S, Cirillo, E, Stewart K., Coort S, Evelo CT, Wesselius A, Zeegers MP, Schols AMW.J. (2019). "Network Analysis of Genome-Wide Association Studies for COPD in the Context of Biological Pathways." American journal of respiratory and critical care medicine.

Mount, S, Houben K, Gosker H, van Beers M, Schuurman L, Franssen F, Janssen D, Schols AMWJ. A randomised-placebo controlled study into the Feasibility and efficacy of working memory training in Chronic Obstructive Pulmonary Disease: the study protocol – submitted

Mount, S., Stewart KFJ, Leerschool AR, Wesselius A, Zeegers MP, Schols AMWJ. Genetic, lifestyle, and environmental factors associated with development of chronic obstructive pulmonary disease – submitted

Mount, S, Wesselius A, van der Grinten C, Savelberg H, Van Geel T, Schaper N, Reijven P, Dagniele P, van Boxtel M, Verhey F, Scham M, Bosma H, Dukers -Muijrers N, Stehouwer C, Schols AMWJ, Zeegers MP. Validation of the Healthy Ageing Phenotype. – in preparation

Curriculum vitae

Sarah Mount was born in North York, Ontario Canada on February 2, 1987. In 2005 she graduated from Jacob Hespeler Secondary School in Cambridge and was admitted to the Nutrition and Nutraceutical Sciences Bachelor programme at the University of Guelph in Guelph Ontario Canada. During this time, she was involved in various extracurricular activities but most notably partook in a bilateral exchange programme with Wageningen University and Research Centre (WUR). Upon graduation she was accepted to the Nutrition and Health Master's programme at the WUR. Here she determined the glycaemic index of palm sugar and made regulatory recommendations for Newtricious BV's leading products. Upon completing this, Sarah won an international scholarship to the University of Reading, Reading England, to investigate the influence of cocoa flavanols on visual function. However, this project yielded negative results, and given further research funding was based on positive results, Sarah decided to end this project and look on elsewhere.

In 2014 Sarah began her PhD trajectory at the department of Respiratory Medicine at Maastricht University. This PhD project was under the supervision of prof. dr. Annemie Schols, prof. dr. Maurice Zeegers and dr. Anke Wesselius. The primary focus of her PhD project was the construction of a Health index to attempt to measure the ageing trajectory. During this project she also investigated a possible adjuvant therapy for COPD rehabilitation as well as the genetic and lifestyle components to COPD risk.

Acknowledgements

To help offset the carbon footprint of this thesis 100 trees were planted.

In this chapter I would like to thank everyone who has made a contribution to this thesis. There are far too many to mention but I would like to begin by thanking my legs for supporting me, my arms for always being by my side, and my fingers, I could always count on them – "Find the humour in difficult situations"- Jeanne Robertson. People tell me this is the only section people actually read, so I have to at least try to be somewhat funny, even if I could never quite show my humour/prankster personality in Dutch very well.

First, I would like to thank my supervisors for giving me the opportunity to do a PhD and come to this wonderful city of Maastricht. Annemie thank you for your constant support, always being available, your constructive feedback, even when we disagreed. Thank you for your patience in my somewhat unconventional PhD approach. I have learned a great deal from you, I will always admire the power and career woman that you are, and for this I thank you. Maurice, thank you for the feedback during the PhD process. Anke, thank you for you for your on-going support, patience and always helpful feedback. Your help with statistics was priceless. I definitely couldn't have finished this PhD without it, or your listening ear through which I vented some frustration, and shared the experience of our first borns.

To the members of my assessment committee, Prof. dr. E.E. Blaak, Prof. dr. L. de Groot, Prof. dr. A. Jansen, Prof. dr. A-H. Maitland van der Zee, and Prof. dr. F-J. van Schooten, thank you for taking the time to review my dissertation.

Thank you to my co-authors in general for your value feedback, time and effort you provided to our articles. Special thanks to Chris Everlo, Susan Coort, Jon Mathers, Elisa Cirillo, José Lara and Luigi Ferrucci for the level of involvement and effort your put into our papers. The results of our collaborations have not only made the resulting papers better, but have also made me a better researcher by allowing me access to data bases, new techniques and topics. Special thanks to Chris van de Grinten for helping me with the Maastricht Study Database, may you rest in peace.

To my fellow mother-PhD'ers, a.k.a crazy people, Emily, Coby, Karin, Kelly and Alexandra. Thank you for reminding me a PhD is a job and not a lifestyle choice. That we can have both, and make our own paths. We can be mothers and still have careers. Thank you for sharing the up's and downs, the frustrations, the laughter and the tears with me, the unique understanding that can only come from someone in this place. Emily, thank you for always being there, just a text away, your sweet gestures made me smile in the most difficult and trying of times. Coby, it was a pleasure working with you and our chats always brightened my day…the offers of vlaai were also of course always appreciated. Karin, thank you for your friendship and being able to share the ups, downs, secrets, and thank you for being there when I needed it most. Kelly, I really enjoyed working with you! I wish we could have done more projects together! We challenge each other and agree to disagree which ultimately results in the best result. Thank

you for your friendship and momming-it with me. Having someone to have your back when things hit the fan is truly priceless; you've even been there with me to help out at my home when I just couldn't do everything on my own. Alexandra, thank you for momming- it with me, and always just being a text away to vent frustrations, or an email away for brainstorming, discussing edits, and of course proof reads.

To my colleagues, thank you for the geziligheid and of course invaluable input during lab meetings and presentations.

Anita and Rosanne, thank you for the jokes, the countless bless yous/gezondheids, the gezeligheids, and coffees. Anita, I'm sorry you didn't make it to Canada but glad you were open to my Canadian-ness and of course watching Ethan for when we thought it was a good idea to try and plaster a room ourselves (don't try this, seriously, don't). Roseanne, thank you for the statistical brainstorming sessions, and problem-solving tips, and of course all the other things mentioned and more. Martijn, you weren't always easy to work with but once we figured out how to work with each other we made a good team. Thanks for your enthusiasm, many coffee breaks and quick chats, even when I was forced to skip lunch breaks etc. due to time commitments of my kids. Thank you to Niki and Ramon for your very useful feedback in our network analysis paper, without your insight it wouldn't have been possible. Chiel thank you for your advice, listening ear and of course valuable input in lab meetings. Harry thank you for your extremely useful power point hacks, Microsoft word, your edits, constructive feedback and ideas, your contributions to my papers took them to the next level. Stephanie and Nanca, you were more than just colleagues but friends. Thank you for the geziligheid, a s I don't know

a better English equivalent, the jokes, the many coffee breaks, and get together' outside of the office. You made my PhD more enjoyable by being an integral part of it.

To Peter (an eye for an I…), Bregtje, Viviënne, Rianne, Nick, Judith, Jules, Ilse, Nicky, Coen, Alex, Irene, Juliette, Lisanne, Cheryl, Wouter, Wouter (lab), Wessel, Juanita, Casper, Marco, Mieke, Renske, and Chantel, I may have missed a few names, if so I am sorry, but thank you for your feedback during the lab meetings, and the creating the unique social dynamic that was PUL. To all the new PUL'ers, welcome, and good luck on your own PhD journey. Pro tip, power point has an align function.

Friends, and family you were irreplaceable during this process. You were there for me for my ups and for my downs, and reminded me sometimes life is hard, sometimes it is easy, but you were there for both and to help pull me out when I was in a slump. To my family thank you for getting me this far, for raising me to be (slightly) competitive, for giving me the drive to always do better and demand the best of myself, and a good dose of stubbornness. You know it never bodes well for the person who says I can't do something. I love you Mom, Dad, Erika and thank you for what you have done to help me to become what I am today.

In no particular order, thank you Alexandria, Steve, Alycia, Charlotte, Jo, Kayla, Jess, Kirby, Mark, Roobini, Vidya, Shari, Simone, Haishan and Andrew. Alexandria and Steve, second parents to my children, thank you for your support, your friendship and being there across time zones, and even letting me practice presentations on you! Alycia, your friendship is

priceless and I can't even begin to put words to describe what your friendship has meant to me in the process, through victories, failures, frustrations and just plain old mom to mom level-headed advice to keep my sanity and survive mom-fatigue and work. Charlotte will never forget meeting you at Vlag, and that's saying something for me. Thanks for guiding me through this final PhD process and of course your on-going friendship. Sorry you didn't like Canada; remind me I owe you some chocolate next time I see you. Jo, my super cool friend who worked at NASA. Thanks for your help from proof reading to expert opinion psychological testing to consultant on being politically correct. You kept me from saying knee jerk reactions and allowed me to vent my frustrations. Thanks for you for being there, no matter where in the world you happened to be. Kayla, thank you for being there during this process, just a text away and always a level headed response. Thank you for helping me plan my wedding in the midst of doing a PhD (because I am that level crazy, I can do it all, get married have kids and renovate a house, why not right?!). Thank you for your priceless friendship doing this PhD process even when you're trying to keep your own crazy life together, and reminding me that I can do this, even when I feel like the world is against me. And of course, the mood lightening pictures of Bowie and Curtsie! Long live the crazy cat ladies. Jess, friend, roomie, travel buddy, animal care taker extraordinaire, thanks for listening to the frustrations and my venting (yes, I know this is a theme I vent…a lot), for sharing jokes, adorable animal pictures and for being there, and of course visiting me here in Maastricht! I hope we van visit again soon! I miss Jess! Haishan, Vidya, Roobini, Kirby, Mark and Simone, thank you for your support, your ideas, your CV checks, your networking, your brainstorming, favours, proof reading and even babysitting. Without you I definitely couldn't have done this. Kirby, Mark and Roobini, thank you for being there when I was horribly sick from pneumonia and recovering from emergency surgery!

To my best friend, my rock, my partner in crime and my loving husband, please take note, I did not forget about the thank you this time :P. Thank you for taking on this crazy southern adventure with me, from planning a wedding, building a home and a family and proving that you're by mind side in sickness, health and in good times and in bad. Thank you for not laughing (too hard) at my endeavours into computer programming, thank you for optimising my programmes to run faster, even if this took out my built-in coffee break. Thank you for helping me through my struggles with R, my nemesis, even learning a new programme language just to give me a hand, and even building a programme to save data when the data set I had been given was almost hopeless. Thank you for teaching me so many things about life, and reminding me what is important through our beautiful children, and more importantly, what isn't. Thank you for your constant encouragement and belief in me (except when it came to the rebricking of the driveway…ha ha) I wouldn't have come this far in life without your, travelled this far, dove into new cultures and languages.