

Genomics of chronic cough unravels neurological pathways

Kayesha Coley^{1*}, Catherine John^{1,4}, Jonas Ghouse^{2,3}, David J. Shepherd¹, Nick Shrine¹, Abril G. Izquierdo¹, Stavroula Kanoni⁵, Emma F. Magavern⁵, Richard Packer^{1,4}, Lorcan McGarvey⁶, Jaclyn A. Smith⁷, Henning Bundgaard^{8,9}, Sisse R. Ostrowski^{9,10}, Christian Erikstrup^{11,12}, Ole B. V. Pedersen^{9,13}, David A. van Heel¹⁴, Genes & Health Research Team⁵, William Hennah^{15,16}, Mikko Marttila¹⁷, Robert C. Free^{4,18}, Edward J. Hollox¹⁹, Louise V. Wain^{1,4}, Martin D. Tobin^{1,4}, Chiara Batini^{1,4*}

¹Department of Population Health Sciences, University of Leicester, Leicester, UK

²Laboratory for Molecular Cardiology, Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Building 9312, Henrik Harpestrengs Vej 4C, 2100 Copenhagen, Denmark

³Laboratory for Molecular Cardiology, Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark

⁴University Hospitals of Leicester NHS Trust, Groby Road, Leicester LE3 9QP, UK

⁵William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

⁶Wellcome-Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK

⁷Division of Immunology, Immunity to Infection and Respiratory Medicine, The University of Manchester, Manchester University NHS Foundation Trust, Manchester, UK

⁸Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

⁹Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

¹⁰Department of Clinical Immunology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

¹¹Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark

¹²Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

¹³Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark

¹⁴Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

¹⁵Orion Pharma, Espoo, Finland

¹⁶Neuroscience Center, HiLIFE, University of Helsinki, Helsinki, Finland

¹⁷Orion Pharma, Nottingham, UK

¹⁸School of Computing and Mathematical Sciences, University of Leicester, Leicester, UK

¹⁹Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

*Correspondence: kc330@leicester.ac.uk (KC) and cb334@leicester.ac.uk (CB)

Abstract

Background

Chronic cough is a symptom of common lung conditions, occurs as a side effect of ACE inhibitors (ACEis), or may be unexplained. Despite chronic cough representing a substantial health burden, its biological mechanisms remain unclear. We hypothesised shared genetic architecture between chronic dry cough and ACEi-induced cough and aimed to identify causal genes underlying both phenotypes.

Methods

We performed multi-ancestry genome-wide association studies (GWAS) of chronic dry cough and ACEi-induced cough, and a multi-trait GWAS of both phenotypes, utilising data from five cohort studies. Chronic dry cough was defined by questionnaire responses, and ACEi-induced cough by treatment switches or clinical diagnosis in electronic health records. We mapped putative causal genes and performed phenome-wide association studies (PheWAS) of associated variants and genetic risk scores (GRS) for these phenotypes to identify pleiotropic effects.

Findings

We found seven novel genetic association signals reaching p -value $<5 \times 10^{-8}$ in the multi-trait or single-trait analyses of chronic dry cough and ACEi-induced cough. The novel variants mapped to 10 novel genes, and we mapped an additional three novel genes to known risk variants, many of which implicating neurological functions (*CTNNA1*, *KCNA10*, *MAPKAP1*, *OR4C12*, *OR4C13*, *SIL1*). The GRS-PheWAS highlighted associations with increased risk of several conditions reported as comorbidities of chronic cough, including fibromyalgia pain, and with spirometry measurements.

Interpretation

Our findings advance the understanding of neuronal dysfunction underlying cough hypersensitivity in chronic dry cough and ACEi-induced cough at the population-level, and the identification of comorbidities associated with genetic predisposition to cough could inform drug target discovery.

Funding

Medical Research Council, Wellcome Trust, National Institute for Health and Care Research, Orion Pharma.

Research in context

Evidence before this study

We searched the National Human Genome Research Institute-European Bioinformatic Institute Catalog of human genome-wide association studies (GWAS) from inception through to 21st May 2024, using the search term “cough” to identify publications which tested association between genetic variants and cough. Additionally, we searched PubMed for English language articles published before 21st May 2024 using the terms “cough” or “ACE inhibitor” combined with “genome-wide association” to find relevant publications. We manually filtered the results from both searches to ensure the cough-related publications related to a dry, unproductive cough. To date, there have been no GWAS of unexplained dry cough, while several for ACE inhibitor (ACEi)-induced cough have identified 11 associated genetic loci at genome-wide significance (p -value $<5 \times 10^{-8}$). These loci have implicated genes involved in neuronal excitability and the bradykinin pathway.

Added value of this study

We present the first multi-ancestry GWAS of chronic dry cough, defined using questionnaire responses, and the largest multi-ancestry GWAS of ACEi-induced cough, characterised by either a medication switch from an ACEi to an angiotensin-II receptor blocker or a clinical diagnosis. By leveraging the clinical and genetic overlap between these two traits, we conducted a multi-trait GWAS to identify novel associated loci. Across the multi-trait and single-trait analyses, we identified seven novel associated loci at genome-wide significance (p -value $<5 \times 10^{-8}$), mapped to 10 genes which have not previously been linked to cough phenotypes.

Implications of all the available evidence

We identified novel genes involved in neurological processes (*CTNNA1*, *KCNA10*, *OR4C12*, *OR4C13*) or implicated in neurological or neurodegenerative conditions (*SIL1* and *MAPKAP1*). The genes we implicate from all genetic associations with cough to date support the role of neurobiological processes underlying dry cough biology, and highlight potential therapeutic targets.

Introduction

Chronic cough is common, impairs quality of life and results in substantial healthcare system burden¹. Despite this, its molecular causes remain unclear and treatments are under-developed². It may occur due to lung diseases, other identifiable conditions (known as ‘treatable traits’) such as ACE inhibitor (ACEi) use, eosinophilic airway inflammation, smoking, obesity, gastro-oesophageal reflux, rhinitis and obstructive sleep apnoea, or remain unexplained with no identifiable comorbidities³⁻⁶. The term cough hypersensitivity syndrome has been used to describe troublesome coughing triggered by low levels of thermal, mechanical or chemical exposure⁷. Neuronal dysregulation has been postulated to underlie cough hypersensitivity, which is thought to be a distinct treatable trait⁸. Recent studies have demonstrated the efficacy of therapies targeting neuronal mechanisms in patients with features consistent with cough hypersensitivity, diagnosed with refractory or unexplained chronic cough^{2,7}.

Genetic studies guide understanding of mechanisms and drug targets⁹ but whilst the genetics of asthma, chronic obstructive pulmonary disease and chronic sputum production have been well-studied¹⁰⁻¹², there has been a lack of genome-wide association studies (GWAS) of chronic cough. In this study, we explore two cough phenotypes: chronic dry cough and ACEi-induced cough, both of which have similar clinical manifestations, including a dry nature and involvement of cough reflex hypersensitivity. There have been no GWAS of chronic dry cough and few for ACEi-induced cough¹³⁻¹⁶, with findings from the latter implicating neuronal excitability and the bradykinin pathway^{13,16}. We hypothesised that there would be common genetic determinants of chronic dry cough and ACEi-induced cough, in which case a multi-trait GWAS increases power to detect shared associated genetic variants¹⁷.

Therefore we: (i) performed the first GWAS of chronic dry cough; (ii) conducted the largest multi-ancestry study of ACEi-induced cough; (iii) tested genetic correlation between chronic dry cough and ACEi-induced cough; (iv) undertook the first multi-trait GWAS of chronic dry cough and ACEi-induced cough; (v) implemented a consensus-based framework to investigate putative causal genes for these traits and (vi) applied phenome-wide association studies (PheWAS) to associated variants across single-trait and multi-trait GWAS, as well as a multi-trait weighted genetic risk score (GRS) to inform understanding of the potential consequences of perturbing pathways involved in chronic cough. Using these approaches, we detected novel genetic associations and identified causal genes and pathways underlying these under-studied cough phenotypes.

Methods

Study populations and phenotype definitions

Our study included individuals from UK Biobank¹⁸, EXCEED¹⁹, Genes & Health²⁰ and Copenhagen Hospital Biobank^{13,21}, and summary statistics from a previously published study conducted in the eMERGE Network^{16,22} (further cohort details provided in **Appendix pp 2**).

We defined ACEi-induced cough using electronic health records (EHRs) linked to UK Biobank¹⁸, EXCEED¹⁹, Genes & Health²⁰ and Copenhagen Hospital Biobank^{13,21} whereby cases switched from an ACEi to an angiotensin-II receptor blocker (ARB) and controls were continuous users of ACEis. The eMERGE Network study defined ACEi-induced cough using a validated algorithm based on prescriptions and the allergy section in EHRs^{16,22} (**Appendix pp 2**).

We defined chronic dry cough in UK Biobank¹⁸ using questionnaire data. Among UK Biobank participants who were neither ACEi-induced cough cases or controls, we defined chronic dry cough cases (non-productive cough on most days) and controls who neither coughed on most days nor brought up sputum on most days (**Appendix pp 2**).

Genome-wide association studies

We performed ancestry-specific GWAS of chronic dry cough and ACEi-induced cough using imputed genomic variants (imputation quality ≥ 0.3 , and minor allele count [MAC] ≥ 10), and age, age², sex, genotyping array and at least 10 principal components of genetic ancestry as covariates. For chronic dry cough, ever-smoking status was an additional covariate. We also utilised summary statistics from the previously published eMERGE Network study¹⁶ (**Appendix pp 2,5-6**).

Using a fixed effects inverse variance-weighted model²³ we undertook a GWAS meta-analysis for chronic dry cough and a GWAS meta-analysis for ACEi-induced cough (**Figure 1; Appendix pp 2**). Then we conducted a multi-trait GWAS¹⁷ by meta-analysis of the chronic dry cough GWAS and ACEi-induced cough GWAS using a fixed effects inverse-variance weighted model²³ (**Figure 1; Appendix pp 2**). To inform the GRS development detailed below, we also performed a European (EUR)-only multi-trait GWAS.

In each GWAS meta-analysis, we defined sentinel variants as those reaching genome-wide significance (p -value $< 5 \times 10^{-8}$) with the lowest p -value within a ± 1 Mb locus. We identified secondary sentinels (p -value $< 5 \times 10^{-8}$) within each locus using conditional analyses in each ancestry group using GCTA^{24,25} (**Appendix pp 3**). We fine-mapped each sentinel locus to calculate the posterior inclusion probability (PIP) for each variant, obtaining sets of variants with a 95% probability of containing a causal variant, referred to as 'credible sets' (**Appendix pp 3**).

We defined single-trait sentinels as independent from multi-trait sentinels based on a linkage disequilibrium (LD) threshold of $r^2 < 0.1$. We also used $r^2 < 0.1$ to define novel multi-trait and single-trait (ACEi-induced cough) sentinels from previously reported associations (p -value

$<5 \times 10^{-8}$) for ACEi-induced cough^{13,15,16}, given the absence of published GWAS of chronic dry cough.

We estimated SNP heritability of chronic dry cough and of ACEi-induced cough and genetic correlation between these traits using LD Score Regression (LDSC)^{26,27} in European (EUR) individuals (**Appendix pp 3**).

Investigating clinical and biological relevance of genetic associations

We mapped genes to all sentinel variants, both novel and previously reported, by integrating the following evidence: (i) nearest gene; (ii) a credible set containing a deleterious or damaging variant with a PIP $\geq 10\%$; (iii) the gene with the highest Polygenic Priority Score (PoPS)²⁸ within $\pm 250\text{kb}$ of the sentinel for the relevant trait (multi-trait, chronic dry cough or ACEi-induced cough); (iv) exome-wide rare variant association (p -value $< 5 \times 10^{-6}$) with the relevant trait in UK Biobank within $\pm 250\text{kb}$ of the sentinel; (v) association with gene expression and (vi) protein expression (**Appendix pp 3-4**). Mapped genes supported by at least one level of evidence were considered putative causal, and we used database (GeneCards²⁹, Open Targets Platform³⁰, Online Mendelian Inheritance in Man³¹ [OMIM], Drug Gene Interaction Database [DGIdb]³²) and literature searches to retrieve information about their biological function and clinical relevance.

We calculated GRS in UK Biobank EUR individuals, comprising all sentinel variants, each weighted by the estimated effect in the EUR-only multi-trait GWAS of chronic dry cough and ACEi-induced cough (outlined above; **Appendix pp 4**). We used DeepPheWAS³³ to test association between sentinels individually (across multiple ancestries) and collectively via the GRS (in EUR only) and 1,939 phenotypes defined in UK Biobank, highlighting associations with a false discovery rate (FDR) < 0.01 . This was complemented by a custom Respiratory PheWAS to query variants against the most powerful GWASs performed across ten clinical respiratory traits (**Appendix pp 4,16**) and identifying associations with a p -value < 0.001 , and by querying Open Targets Genetics³⁴ to assess whether sentinels were in LD ($r^2 > 0.8$) with genome-wide significant lead GWAS variants for other traits.

Sensitivity analysis

We performed a sensitivity analysis in UK Biobank excluding individuals with asthma (defined by DeepPheWAS³³ using both self-report and EHR data). For our sentinel variants, we followed the primary analysis protocol (outlined above; **Appendix pp 2,5-6**) and performed ancestry-specific association testing with chronic dry cough and ACEi-induced cough, excluding asthma cases. Using an inverse-variance weighted fixed effects model, we conducted separate meta-analyses for each trait, followed by a combined multi-trait meta-analysis of both traits.

Role of the funding source

Orion Pharma supported this work through a funded research collaboration with the University of Leicester, through which DeepPheWAS was developed. We used DeepPheWAS for the PheWAS analyses presented in this paper and co-authors William Hennah and Mikko Marttila are salaried employees of Orion Pharma.

Results

The multi-trait GWAS of chronic dry cough and ACEi-induced cough included 29,158 cases and 151,222 controls, comprising 7,635 cases and 88,653 controls for chronic dry cough, and 21,523 cases and 62,569 controls for ACEi-induced cough (**Figure 1; Appendix pp 6**). SNP heritability for chronic dry cough was 3.19% (SE = 0.53%) and for ACEi-induced cough was 5.30% (SE 1.4%) and there was a strong genetic correlation between chronic dry cough and ACEi-induced cough ($r_g = 0.56$, SE = 0.15, p -value = 0.0001). Across the multi-trait and single-trait analyses, we defined 14 independent signals, identified by sentinel variants reaching p -value $< 5 \times 10^{-8}$ (**Appendix pp 7-8**). These signals mapped to a total of 19 genes through our comprehensive variant-to-gene mapping approach (**Figure 3; Appendix pp 9-12,25**), and were subsequently explored for gene-drug interactions (**Appendix pp 13**).

We found seven novel genetic association signals in the multi-trait or single-trait analyses of chronic dry cough and ACEi-induced cough (**Table 1**, see **Appendix pp 7-8** for further details). In the multi-trait GWAS, four novel signals mapped to six genes, five of which (*CTNNA1*, *OR4C12*, *OR4C13*, *SIL1* and *VMA21*) have not been previously implicated in any cough GWAS (**Figure 2; Figure 3**). The remaining gene, *L3MBTL4*, was previously implicated by a common variant (rs8097200) associated with ACEi-induced cough¹³, and independent from our sentinel, rs78598167 ($r^2 = 0.003$).

The single-trait analyses identified three novel signals mapping to 5 novel genes (**Table 1**). While the sentinel variant found for chronic dry cough (rs141733360), and its mapped gene (*CPEB2*), have not been previously described in the literature as associated with any cough phenotype, the only phenotypes we observed associated with rs141733360 in the PheWAS were cough-related (**Appendix pp 14**). The two sentinel variants for ACEi-induced cough mapped to *ATP23* and *CYP27B1* (rs11172406), and *ALX1* and *RASSF9* (rs35336617), none of which have been identified before for any cough phenotype. Our PheWAS showed that the allele associated with increased ACEi-induced cough (allele A of rs11172406), mapping to *ATP23* and *CYP27B1*, was associated with reduced urea and urate levels. This finding was supported by the Open Targets query for its proxy rs10877067 (in high LD with rs11172406, $r^2 = 0.88$); allele T at this variant which increased risk of ACEi-induced cough was also associated with reduced urate levels (**Appendix pp 14-15**).

We also identified seven previously reported signals for ACEi-induced cough, six in the multi-trait analysis and one in the ACEi-induced cough single-trait analysis. Collectively, these mapped to nine genes (*KCNA10*, *KCNIP4*, *L3MBTL4*, *MAPKAP1*, *NTSR1*, *PREP*, *RBM15*, *SLCO4A1* and *SRBD1*), three of which (*KCNA10*, *MAPKAP1* and *RBM15*) were never implicated before for any cough phenotype (**Appendix pp 7**). The PheWAS showed that the signal mapping to *MAPKAP1* was associated with increased risk of asthma¹² (p -value = 3.70×10^{-4}) [**Appendix pp 16-17**], while the signal mapping to *NTSR1* and *SLCO4A1* was associated with both risk of asthma¹² (p -value = 1.70×10^{-5}) [**Appendix pp 16-17**] and cough-related phenotypes (FDR $< 1\%$, **Appendix pp 14**) with the same direction of effect. However, the effect size estimates of all sentinel variants were not systematically or substantially attenuated after exclusion of asthma cases (**Appendix pp 18**).

We identified the gene product of *KCNA10* as a target for several potassium channel blockers, including fampridine, which has been approved for multiple sclerosis treatment, as

well as others currently undergoing clinical trials for other indications including stroke and spinal cord injury (**Appendix pp 13**). Reminertant, a drug targeting the protein encoded by *NTSR1* is also undergoing trials for small cell lung carcinoma (**Appendix pp 13**).

Five of the seven sentinels previously reported for ACEi-induced cough¹³ showed association (p -value < 0.05) with chronic dry cough, more than expected by chance (binomial test p -value = 6×10^{-6}). These mapped to the *SRBD1*, *KCNIP4*, *PREP*, *SCAI* and *NTSR1/SLCO4A1* genes¹³, all five variants showing a consistent direction of effect between chronic dry cough and ACEi-induced cough and a larger effect size estimate for chronic dry cough than for cough on most days¹³. The association with chronic dry cough met a Bonferroni threshold for the sentinels at the *SRBD1* (rs1544730), *PREP* (rs7761208), and *NTSR1/SLCO4A1* (rs6062847) loci.

We developed a GRS for chronic cough based on the 14 sentinel variants weighted by the multi-trait analysis to maximise power for the PheWAS. Genetically predicted chronic cough was associated with quantitative phenotypes including lower forced expiratory volume in one second and lower forced vital capacity, lower fat free mass, reduced hand grip strength and lower urate levels (**Appendix pp 19-24**). Furthermore, an increase in the GRS was associated with disease risk, such as an increased risk of breast cancer and osteoporosis, and reduced risk of gout, diabetes and ischaemic heart disease. Most notably, an increase in the GRS was associated with increased risk of migraine, irritable bowel syndrome (IBS), fibromyalgia pain, hypothyroidism and urinary incontinence – all of which have been reported as comorbidities accompanying chronic cough³⁵⁻³⁸ (**Appendix pp 19-24,26**).

Discussion

Chronic cough is a common symptom with prevalence estimates varying from 1-2% up to 10%^{4,39}, and imposes a substantial health burden¹. Despite this, its biological mechanisms are not fully understood. We have utilised GWAS approaches to identify associated genetic variants and implicate relevant genes, shedding light on the molecular basis of chronic dry cough and ACEi-induced cough, and providing genetic evidence for future therapeutic developments. Specifically, we undertook the first GWAS of chronic dry cough and the largest multi-ancestry GWAS of ACEi-induced cough to date, and having demonstrated genetic overlap between chronic dry cough and ACEi-induced cough, we performed a multi-trait GWAS, thereby maximising power to discover novel variants associated with both traits. Further, we showed that collectively, cough-associated variants were associated with raised risk of conditions such as migraine, IBS, fibromyalgia pain, hypothyroidism, urinary incontinence, breast cancer, and osteoporosis, providing further insights as to the potential consequences of modulation of pathways involved in chronic cough.

We found seven novel sentinel variants associated with these cough phenotypes which reached p -value $<5 \times 10^{-8}$ in the multi-trait or single-trait analyses and highlight the relevance for chronic dry cough of five further sentinels previously reported for ACEi-induced cough. The seven novel sentinels mapped to 10 genes not previously implicated in cough phenotypes (*SIL1*, *CTNNA1*, *OR4C12*, *OR4C13*, *VMA21*, *CPEB2*, *ATP23*, *CYP27B1*, *ALX1*, *RASSF9*), and through a comprehensive variant-to-gene mapping strategy we additionally mapped known loci to a further three novel genes (*MAPKAP1*, *KCNA10*, *RMB15*). Descriptions of novel (and previously reported) genes are provided in **Appendix pp 25**.

Collectively, the novel sentinels and genes, alongside those previously reported, enrich our understanding of the biological mechanisms underlying chronic dry cough and ACEi-induced cough. Many of the novel putative causal genes encode proteins with neurological functions. *CTNNA1* is a member of the catenin protein family which is involved in cell-cell adhesion²⁹, synapse morphogenesis and plasticity⁴⁰. *KCNA10* encodes a subunit of potassium ion channels (as does *KCNIP4*¹³ highlighted below) which have a role in neuronal excitability, neurotransmitter release and smooth muscle contraction^{29,30}, and is a target for an approved drug for symptom management in multiple sclerosis. *MAPKAP1* is a subunit of the mTOR complex 2, a component of the mTOR signalling pathway vital for growth and metabolism^{29,30,41}. *MAPKAP1* has been identified in GWAS of pain intensity⁴², and dysregulation of the mTOR signalling pathway has been implicated in various neurodegenerative and neuropsychiatric disorders, including epilepsy⁴³. *OR4C12* and *OR4C13* are G-protein coupled receptors predominantly expressed in nasal epithelium olfactory sensory neurons and involved in neural response to odorant stimuli^{29,30}. *SIL1* is a glycoprotein involved in protein translocation and folding and is associated with Marinesco-Sjogren syndrome³¹ which is characterised by a number of neurological symptoms such as cerebellar ataxia and intellectual disability⁴⁴.

Further, we provide evidence that loci, genes and pathways previously implicated in ACEi-induced cough have a role in chronic dry cough, and may provide insights into potential therapeutic targets^{13,16,45}. The genes previously implicated for ACEi-induced cough also highlight the importance of neurological processes. In particular, *PREP* is involved in the

mediation of neuropeptide activity, including the metabolism of bradykinin^{46,47} which has been shown to sensitise airway sensory nerves and induce cough reflex hypersensitivity in an animal model via the activation of airway C-fibres⁴⁸. *KCNIP4* modulates neuronal excitability²⁹, while NTSR1 mediates the activity of neurotensin, a pro-inflammatory peptide involved in the modulation of smooth muscle contraction⁴⁹, and also upregulates the secretion of GABA³⁰. Animal models and controlled trials have demonstrated that centrally acting agonists of GABA_B receptors inhibit the cough reflex; with baclofen being reported to suppress ACEi-induced cough^{50,51}. Additionally, animal studies have provided evidence for the role of neuronal potassium channels in smooth muscle reactivity and cough reflex modulation⁵².

We found that GRSs for our cough phenotypes were associated with increased risk of several recognized comorbidities of chronic cough, including fibromyalgia pain. The neurobiological mechanisms of chronic cough and chronic pain exhibit similarities, with central and peripheral sensitization contributing to their underlying symptoms, and afferent nerve fibers involved in both conditions expressing common receptors⁵³. P2X3 receptors are involved in nerve fiber sensitization in chronic pain⁵⁴, and P2X3 receptor antagonists have shown antitussive efficacy in refractory and unexplained chronic cough patients in clinical trials⁵⁵.

By leveraging questionnaire responses from UK Biobank to define chronic dry cough and primary care EHR prescribing data to characterise ACEi-induced cough, we conducted the most powerful GWAS of chronic cough to date. This phenotyping approach allowed us to maximise the use of relevant phenotypic data available and overcome the limitation that drug response and adverse reaction-related traits are rarely specifically recorded in study populations, facilitating the study of ACEi-induced cough in large sample sizes across multiple cohorts. Another strength of this study is the implementation of a detailed variant-to-gene mapping framework to implicate putative causal genes and biological mechanisms. Further, we utilised recently developed techniques to improve power for understanding the potential consequences of pathway modulation through GRS-informed PheWAS.

We also note limitations of our study. Although we included individuals of diverse ancestries, the sample sizes for non-European groups remained modest despite our efforts to maximise data availability by utilising both EHRs and questionnaire responses. This bias presents an important issue for many GWASs⁵⁶ and highlights the importance of initiatives with a recruitment focus on underrepresented populations to improve generalisability of findings across populations⁵⁷. Additionally, as with any case-control study, misclassification of cases and controls is possible, perhaps more so using an ACEi to ARB switch to characterise cases. However, such misclassification would tend to reduce effect size estimates towards the null and potentially miss some true associations, and an ACEi to ARB switch to capture ACEi-induced cough has been validated through manual interrogation of patient histories¹⁵, consistency of genetic associations^{13,16} and polygenic risk score predictions¹³. We note that variable terms are used to describe and categorise chronic cough, including cough hypersensitivity syndrome, refractory chronic cough and unexplained chronic cough⁵⁸. We did not attempt to distinguish these conditions, which would generally require specialist clinical assessment given that cough is poorly coded in EHRs⁶ and not widely characterised in large cohort questionnaires. Such studies will require further development of registries of

well-characterised patients with samples and consent for genetic studies, such as NEuroCOUGH².

Our findings expand on the established concept of cough hypersensitivity due to neuronal dysfunction, which in specific cases can be identified as a treatable trait, by showing that this mechanism applies broadly to chronic dry cough at the population-level, and to ACEi-induced cough. Further, identifying the association between genetic predisposition to cough with various known comorbidities, including chronic pain, could further advance our understanding of involved biological processes and should be considered in drug discovery and development efforts.

Tables

Table 1. Novel sentinel variants (p -value $<5 \times 10^{-8}$) identified in the multi-trait GWAS of chronic dry cough and ACEi-induced cough, and additional novel sentinels from each single-trait GWAS. Abbreviations: CI, confidence interval; OR, odds ratio.

Sentinel ID	Position (hg19)	Trait	Effect allele	Effect allele frequency	OR [95% CI]	p -value	Annotation	Mapped gene(s)
5:138404361:A:AAAG	5:138404361	Multi-trait	A	0.326	1.08 [1.03, 1.13]	2.62×10^{-8}	Intronic	<i>CTNNA1</i> ; <i>SIL1</i>
rs112658458	11:49790181	Multi-trait	A	0.967	1.19 [1.14, 1.32]	8.31×10^{-9}	Intronic	<i>OR4C12</i> ; <i>OR4C13</i>
rs78598167	18:6110865	Multi-trait	T	0.023	1.33 [1.31, 1.56]	8.36×10^{-15}	Intronic	<i>L3MBTL4</i>
rs5924943	X:150594104	Multi-trait	T	0.469	1.07 [1.07, 1.12]	3.76×10^{-16}	Intergenic	<i>VMA21</i>
rs141733360	4:14727977	Chronic dry cough	A	0.009	0.51 [0.41, 0.64]	4.20×10^{-9}	ncRNA intronic	<i>CPEB2</i>
rs11172406	12:58366606	ACEi-induced cough	A	0.361	1.11 [1.07, 1.15]	1.73×10^{-8}	Intergenic	<i>ATP23</i> ; <i>CYP27B1</i>
rs35336617	12:85913015	ACEi-induced cough	T	0.703	0.87 [0.83, 0.91]	2.71×10^{-9}	Intergenic	<i>ALX1</i> ; <i>RASSF9</i>

Figures

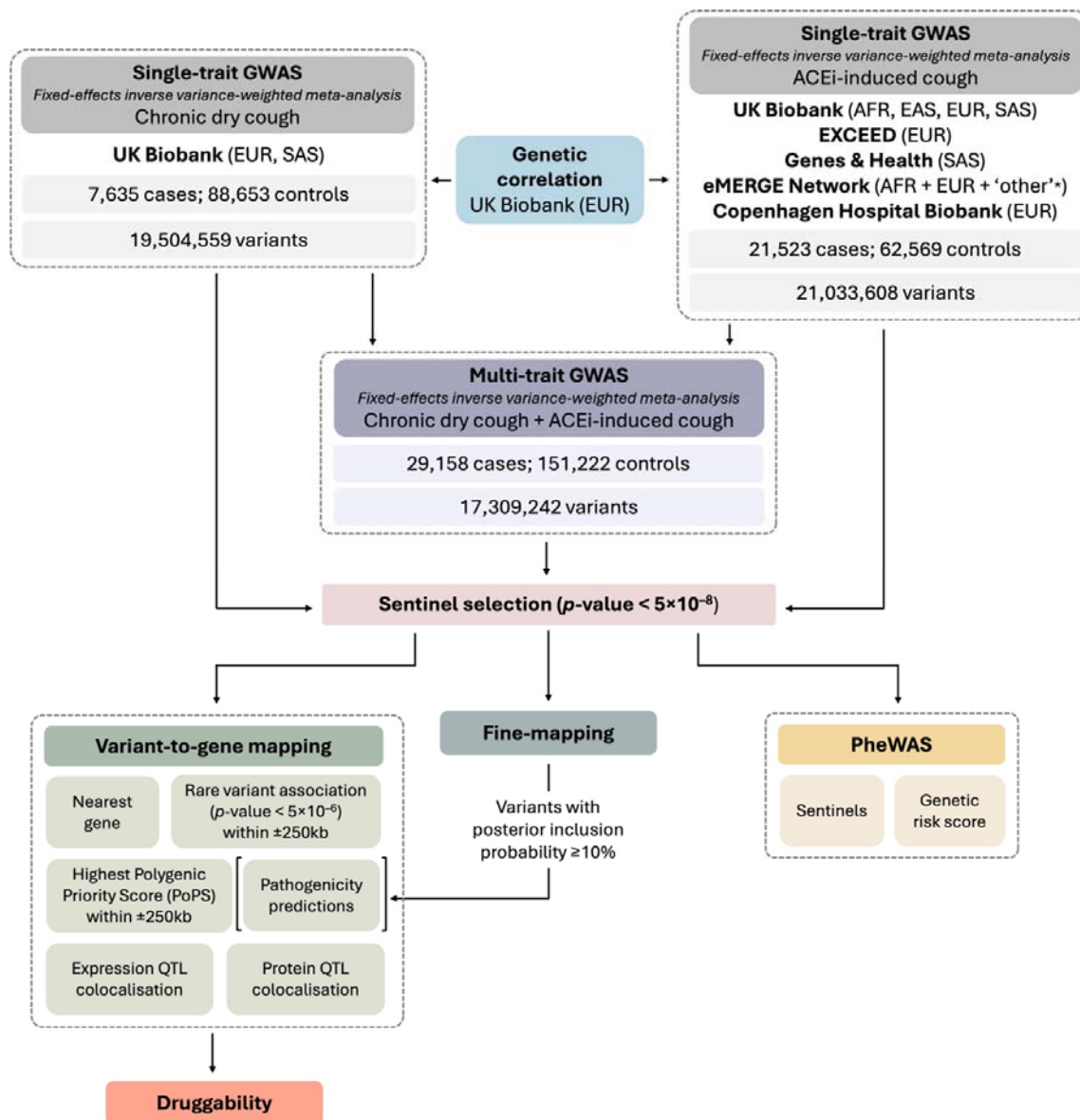


Figure 1. Study design. Abbreviations: ACEi, ACE inhibitor; AFR, African; EAS, East Asian; EUR, European; PheWAS, phenome-wide association study; SAS, South Asian; QTL, quantitative trait loci. *Defined by study authors¹⁶.

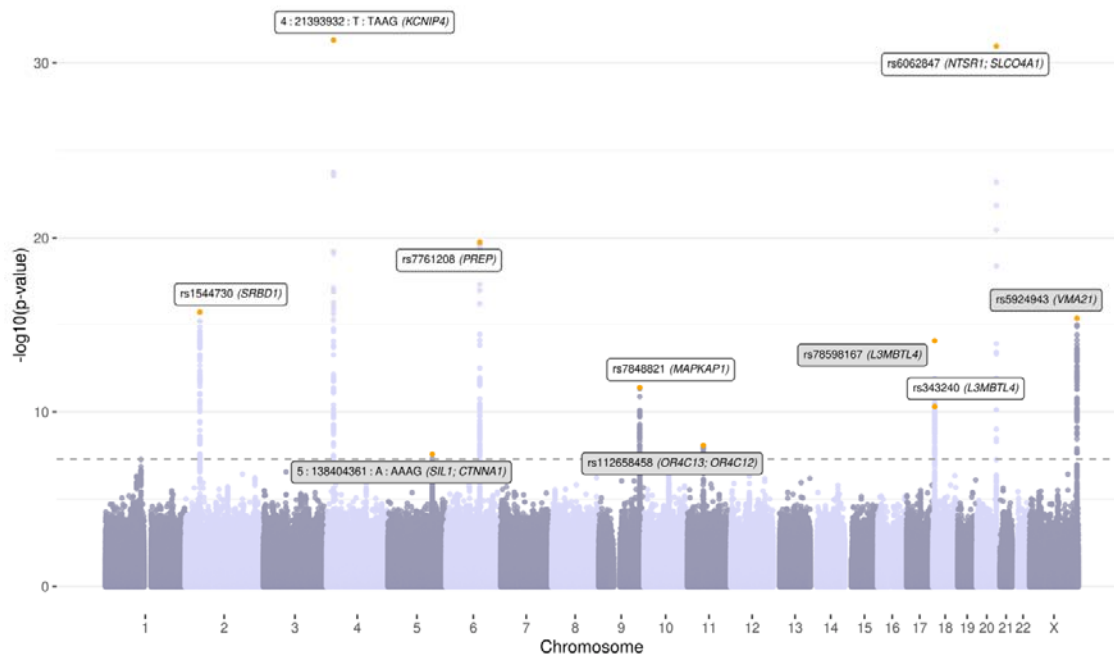


Figure 2. Manhattan plot for the multi-trait GWAS of chronic dry cough and ACEi-induced cough. The dotted grey line represents the genome-wide significance threshold (p -value $<5 \times 10^{-8}$). Sentinels are highlighted orange and mapped genes are labelled; novel sentinels are highlighted grey.

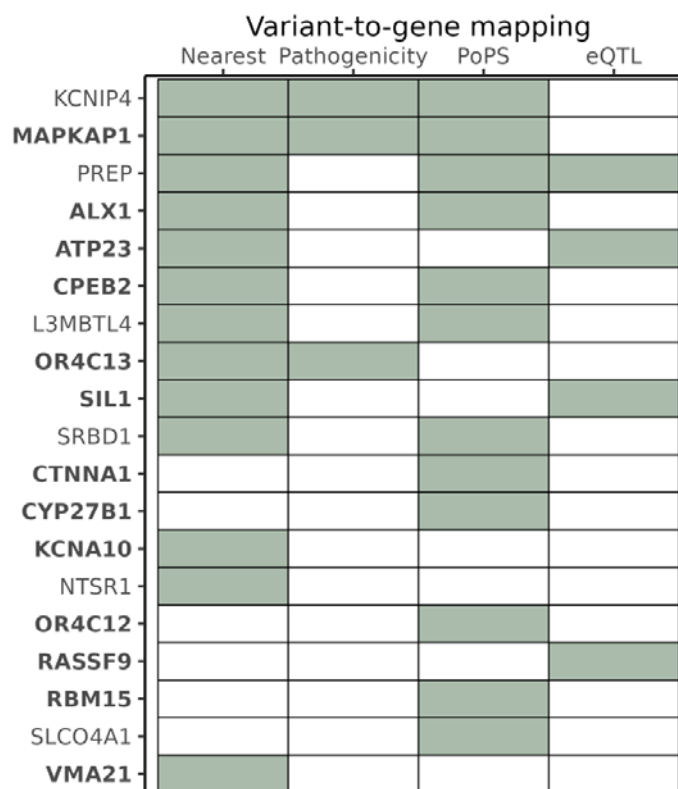


Figure 3. Genes identified by variant-to-gene mapping. Filled tiles indicate genes identified by specific analysis. Protein quantitative trait loci (pQTL) and rare variant columns have been omitted as these did not lead to the mapping of any genes. Novel genes are highlighted bold. *Abbreviations:* eQTL, expression quantitative trait loci; PoPS, Polygenic Priority Score; pQTL, protein quantitative trait loci.

References

1. Bali V, Adriano A, Byrne A, Akers KG, Frederickson A, Schelfhout J. Understanding the economic burden of chronic cough: a systematic literature review. *BMC Pulm Med*. Oct 31 2023;23(1):416. doi:10.1186/s12890-023-02709-9
2. McGarvey L, Dupont L, Birring SS, et al. New understanding in the treatment of cough (NEUROCOUGH) ERS Clinical Research Collaboration: improving care and treatment for patients with cough. *Eur Respir J*. May 2019;53(5)doi:10.1183/13993003.00787-2019
3. Arinze JT, de Roos EW, Karimi L, Verhamme KMC, Stricker BH, Brusselle GG. Prevalence and incidence of, and risk factors for chronic cough in the adult population: the Rotterdam Study. *ERJ Open Res*. Apr 2020;6(2)doi:10.1183/23120541.00300-2019
4. Arinze JT, van der Veer T, Bos D, Stricker B, Verhamme KMC, Brusselle G. Epidemiology of unexplained chronic cough in adults: a population-based study. *ERJ Open Res*. May 2023;9(3)doi:10.1183/23120541.00739-2022
5. Holden SE, Morice A, Birring SS, et al. Cough presentation in primary care and the identification of chronic cough: a need for diagnostic clarity? *Curr Med Res Opin*. Jan 2020;36(1):139-150. doi:10.1080/03007995.2019.1673716
6. Hull JH, Langerman H, Ul-Haq Z, Kamalati T, Lucas A, Levy ML. Burden and impact of chronic cough in UK primary care: a dataset analysis. *BMJ Open*. Dec 17 2021;11(12):e054832. doi:10.1136/bmjopen-2021-054832
7. Parker SM, Smith JA, Birring SS, et al. British Thoracic Society Clinical Statement on chronic cough in adults. *Thorax*. Dec 6 2023;78(Suppl 6):s3-s19. doi:10.1136/thorax-2023-220592
8. Pavord I. New statement and new hope for patients with refractory chronic cough. *Thorax*. Dec 6 2023;78(Suppl 6):s1-s2. doi:10.1136/thorax-2023-220666
9. Minikel EV, Painter JL, Dong CC, Nelson MR. Refining the impact of genetic evidence on clinical success. *Nature*. May 2024;629(8012):624-629. doi:10.1038/s41586-024-07316-0
10. Packer RJ, Shrine N, Hall R, et al. Genome-wide association study of chronic sputum production implicates loci involved in mucus production and infection. *Eur Respir J*. Jun 2023;61(6)doi:10.1183/13993003.01667-2022
11. Shrine N, Izquierdo AG, Chen J, et al. Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk. *Nat Genet*. Mar 2023;55(3):410-422. doi:10.1038/s41588-023-01314-0
12. Tsuo K, Zhou W, Wang Y, et al. Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity. *Cell Genom*. Dec 14 2022;2(12):100212. doi:10.1016/j.xgen.2022.100212
13. Ghouse J, Tragante V, Muhammad A, et al. Polygenic risk score for ACE-inhibitor-associated cough based on the discovery of new genetic loci. *Eur Heart J*. Dec 1 2022;43(45):4707-4718. doi:10.1093/eurheartj/ehac322
14. Hallberg P, Persson M, Axelsson T, et al. Genetic variants associated with angiotensin-converting enzyme inhibitor-induced cough: a genome-wide association study in a Swedish population. *Pharmacogenomics*. Feb 2017;18(3):201-213. doi:10.2217/pgs-2016-0184

15. Mahmoudpour SH, Veluchamy A, Siddiqui MK, et al. Meta-analysis of genome-wide association studies on the intolerance of angiotensin-converting enzyme inhibitors. *Pharmacogenet Genomics*. Mar 2017;27(3):112-119. doi:10.1097/FPC.0000000000000264
16. Mosley JD, Shaffer CM, Van Driest SL, et al. A genome-wide association study identifies variants in KCNIP4 associated with ACE inhibitor-induced cough. *Pharmacogenomics J*. Jun 2016;16(3):231-7. doi:10.1038/tpj.2015.51
17. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet*. Jul 2013;14(7):483-95. doi:10.1038/nrg3461
18. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. Oct 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
19. John C, Reeve NF, Free RC, et al. Cohort profile: Extended Cohort for E-health, Environment and DNA (EXCEED). *Int J Epidemiol*. Oct 1 2019;48(5):1734. doi:10.1093/ije/dyz175
20. Finer S, Martin HC, Khan A, et al. Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int J Epidemiol*. Feb 1 2020;49(1):20-21i. doi:10.1093/ije/dyz174
21. Sorensen E, Christiansen L, Wilkowski B, et al. Data Resource Profile: The Copenhagen Hospital Biobank (CHB). *Int J Epidemiol*. Jul 9 2021;50(3):719-720e. doi:10.1093/ije/dyaa157
22. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med*. Oct 2013;15(10):761-71. doi:10.1038/gim.2013.72
23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. Sep 1 2010;26(17):2190-1. doi:10.1093/bioinformatics/btq340
24. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. Mar 18 2012;44(4):369-75, S1-3. doi:10.1038/ng.2213
25. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. Jan 7 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011
26. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. Nov 2015;47(11):1236-41. doi:10.1038/ng.3406
27. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. Mar 2015;47(3):291-5. doi:10.1038/ng.3211
28. Weeks EM, Ulirsch JC, Cheng NY, et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat Genet*. Jul 13 2023;55(8):1267-1276. doi:10.1038/s41588-023-01443-6
29. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics*. Jun 20 2016;54:1 30 1-1 30 33. doi:10.1002/cpbi.5
30. Ochoa D, Hercules A, Carmona M, et al. The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acids Res*. Jan 6 2023;51(D1):D1353-D1359. doi:10.1093/nar/gkac1046

31. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* Jan 1 2005;33(Database issue):D514-7. doi:10.1093/nar/gki033
32. Freshour SL, Kiwala S, Cotto KC, et al. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsourcing efforts. *Nucleic Acids Res.* Jan 8 2021;49(D1):D1144-D1151. doi:10.1093/nar/gkaa1084
33. Packer RJ, Williams AT, Hennah W, et al. DeepPheWAS: an R package for phenotype generation and association analysis for phenome-wide association studies. *Bioinformatics.* Apr 3 2023;39(4)doi:10.1093/bioinformatics/btad073
34. Ghousaini M, Mountjoy E, Carmona M, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* Jan 8 2021;49(D1):D1311-D1320. doi:10.1093/nar/gkaa840
35. Ford AC, Forman D, Moayyedi P, Morice AH. Cough in the community: a cross sectional survey and the relationship to gastrointestinal symptoms. *Thorax.* Nov 2006;61(11):975-9. doi:10.1136/thx.2006.060087
36. Irwin RS, Dudiki N, French CL, Panel CEC. Life-Threatening and Non-Life-Threatening Complications Associated With Coughing: A Scoping Review. *Chest.* Nov 2020;158(5):2058-2073. doi:10.1016/j.chest.2020.06.012
37. Arinze JT, Verhamme KMC, Luik AI, Stricker B, van Meurs JBJ, Brusselle GG. The interrelatedness of chronic cough and chronic pain. *Eur Respir J.* May 2021;57(5)doi:10.1183/13993003.02651-2020
38. Birring SS, Patel RB, Parker D, et al. Airway function and markers of airway inflammation in patients with treated hypothyroidism. *Thorax.* Mar 2005;60(3):249-53. doi:10.1136/thx.2004.034900
39. Morice AH, Jakes AD, Faruqi S, et al. A worldwide survey of chronic cough: a manifestation of enhanced somatosensory response. *Eur Respir J.* Nov 2014;44(5):1149-55. doi:10.1183/09031936.00217813
40. Arikath J, Reichardt LF. Cadherins and catenins at synapses: roles in synaptogenesis and synaptic plasticity. *Trends Neurosci.* Sep 2008;31(9):487-94. doi:10.1016/j.tins.2008.07.001
41. Takei N, Nawa H. mTOR signaling and its roles in normal and abnormal brain development. *Front Mol Neurosci.* 2014;7:28. doi:10.3389/fnmol.2014.00028
42. Toikumo S, Vickers-Smith R, Jinwala Z, et al. A multi-ancestry genetic study of pain intensity in 598,339 veterans. *Nat Med.* Apr 2024;30(4):1075-1084. doi:10.1038/s41591-024-02839-5
43. Lipton JO, Sahin M. The neurology of mTOR. *Neuron.* Oct 22 2014;84(2):275-91. doi:10.1016/j.neuron.2014.09.034
44. Krieger M, Roos A, Stendel C, et al. SIL1 mutations and clinical spectrum in patients with Marinesco-Sjogren syndrome. *Brain.* Dec 2013;136(Pt 12):3634-44. doi:10.1093/brain/awt283
45. Dicipinigitis PV. Angiotensin-converting enzyme inhibitor-induced cough: ACCP evidence-based clinical practice guidelines. *Chest.* Jan 2006;129(1 Suppl):169S-173S. doi:10.1378/chest.129.1_suppl.169S
46. Serfozo P, Wysocki J, Gulua G, et al. Prolyl Oligopeptidase-Dependent Angiotensin II Conversion to Angiotensin-(1-7) in the circulation. *Hypertension.* Jan 2020;75(1):173-182. doi:10.1161/HYPERTENSIONAHA.119.14071

47. Fox AJ, Lalloo UG, Belvisi MG, Bernareggi M, Chung KF, Barnes PJ. Bradykinin-evoked sensitization of airway sensory nerves: a mechanism for ACE-inhibitor cough. *Nat Med*. Jul 1996;2(7):814-7. doi:10.1038/nm0796-814
48. Hewitt MM, Adams G, Jr., Mazzone SB, Mori N, Yu L, Canning BJ. Pharmacology of Bradykinin-Evoked Coughing in Guinea Pigs. *J Pharmacol Exp Ther*. Jun 2016;357(3):620-8. doi:10.1124/jpet.115.230383
49. Djokic TD, Dusser DJ, Borson DB, Nadel JA. Neutral endopeptidase modulates neurotensin-induced airway contraction. *J Appl Physiol (1985)*. May 1989;66(5):2338-43. doi:10.1152/jappl.1989.66.5.2338
50. Dicipinigaitis PV. Use of baclofen to suppress cough induced by angiotensin-converting enzyme inhibitors. *Ann Pharmacother*. Nov 1996;30(11):1242-5. doi:10.1177/106002809603001106
51. Badri H, C LG, Denton D, et al. Effect of centrally and peripherally acting GABA(B) agonism on the healthy human cough reflex. *Pulm Pharmacol Ther*. Dec 2021;71:102079. doi:10.1016/j.pupt.2021.102079
52. Sutovska M, Nosalova G, Franova S. The role of potassium ion channels in cough and other reflexes of the airways. *J Physiol Pharmacol*. Nov 2007;58 Suppl 5(Pt 2):673-83.
53. O'Neill J, McMahon SB, Udem BJ. Chronic cough and pain: Janus faces in sensory neurobiology? *Pulm Pharmacol Ther*. Oct 2013;26(5):476-85. doi:10.1016/j.pupt.2013.06.010
54. Bernier LP, Ase AR, Seguela P. P2X receptor channels in chronic pain pathways. *Br J Pharmacol*. Jun 2018;175(12):2219-2230. doi:10.1111/bph.13957
55. McGarvey LP, Birring SS, Morice AH, et al. Efficacy and safety of gefapixant, a P2X(3) receptor antagonist, in refractory chronic cough and unexplained chronic cough (COUGH-1 and COUGH-2): results from two double-blind, randomised, parallel-group, placebo-controlled, phase 3 trials. *Lancet*. Mar 5 2022;399(10328):909-923. doi:10.1016/S0140-6736(21)02348-5
56. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell*. Mar 21 2019;177(1):26-31. doi:10.1016/j.cell.2019.02.048
57. Tobin MD, Izquierdo AG. Improving ethnic diversity in respiratory genomics research. *Eur Respir J*. Oct 2021;58(4)doi:10.1183/13993003.01615-2021
58. McGarvey L, Gibson PG. What Is Chronic Cough? Terminology. *J Allergy Clin Immunol Pract*. Jul-Aug 2019;7(6):1711-1714. doi:10.1016/j.jaip.2019.04.012

Data sharing

Access to UK Biobank (<https://www.ukbiobank.ac.uk>), EXCEED (<https://exceed.org.uk/>), Genes & Health (<https://www.genesandhealth.org/>) and Copenhagen Hospital Biobank individual-level data are available to approved researchers upon application or data access request. Genome-wide summary statistics from the single-trait and multi-trait analyses will be made publicly available via the EMBL-EBI GWAS Catalog. The ACEi-induced cough phenotyping algorithm is publicly available (<https://doi.org/10.5281/zenodo.6780065>), and scripts used to run additional analyses are available upon request.

Funding

This study was supported by the following: UKRI (MRC) Innovation Fellowship at Health Data Research UK grant number MR/S003762/1 to CB; Wellcome Trust Investigator Award (WT202849/Z/16/Z); Wellcome Trust Discovery Award (WT225221/Z/22/Z), NIHR Senior Investigator Award (NIHR201371) to MDT and the NIHR Leicester Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. EFM is funded by Barts Charity and Barts NIHR BRC. Orion Pharma supported this work through a funded research collaboration with the University of Leicester, through which DeepPheWAS was developed. We used DeepPheWAS for the PheWAS analyses presented in this paper and co-authors William Hennah and Mikko Marttila are salaried employees of Orion Pharma.

Declaration of Interests

CJ, RP, LVW and MDT report funding from Orion Pharma within the scope of the submitted work. LVW has held research grants from GlaxoSmithKline (as principal investigator) unrelated to current work, and reports consultancy for Galapagos. WH and MM are salaried employees of Orion Pharma. MDT has research collaborations with GlaxoSmithKline unrelated to the current work. CE has received unrestricted research grants from Novo Nordisk and Abbott Diagnostics; no personal fees.

Acknowledgements

We thank all volunteers participating in the EXCEED Study, UK Biobank, Genes & Health, Copenhagen Hospital Biobank and the eMERGE Network and who have made this project possible. We also thank authors of the eMERGE Network study for sharing GWAS summary statistics. This study used the ALICE High Performance Computing Facilities at the University of Leicester.

Copenhagen Hospital Biobank: The analysis was performed under the Copenhagen Hospital Biobank Cardiovascular Disease Cohort (CHB-CVDC), approval number NVK-1708829, P-2019-93.

EXCEED: EXCEED has been supported by the University of Leicester, the National Institute for Health and Care Research Leicester Respiratory Biomedical Research Centre, the Wellcome Trust (WT 202849), and Medical Research Council grants G0902313 and

UKRI_PC_19004, the latter through the UK Research and Innovation Industrial Strategy Challenge Fund, delivered through Health Data Research UK. EXCEED has also been supported by Cohort Access fees from studies funded by the Medical Research Council (MRC), Biotechnology and Biological Sciences Research Council, National Institute for Health and Care Research, the UK Space Agency, and GlaxoSmithKline. EXCEED received ethical approval from the Leicester Central Research Ethics Committee (13/EM/0226). Substantial amendments have been approved by the same Research Ethics Committee for the collection of new data relating to the COVID-19 pandemic, including the COVID-19 questionnaires and antibody testing.

Genes and Health: Genes & Health is/has recently been core-funded by Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017, MR/X009777/1, MR/X009920/1), Higher Education Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London substantive site), and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). Genes & Health is/has recently been funded by Alnylam Pharmaceuticals, Genomics PLC; and a Life Sciences Industry Consortium of Astra Zeneca PLC, Bristol-Myers Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc, Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc, Takeda Development Centre Americas Inc. We thank Social Action for Health, Centre of The Cell, members of our Community Advisory Group, and staff who have recruited and collected data from volunteers. We thank the NIHR National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry Centre (King's College London), Wellcome Sanger Institute, and Broad Institute for sample processing, genotyping, sequencing and variant annotation. We thank: Barts Health NHS Trust, NHS Clinical Commissioning Groups (City and Hackney, Waltham Forest, Tower Hamlets, Newham, Redbridge, Havering, Barking and Dagenham), East London NHS Foundation Trust, Bradford Teaching Hospitals NHS Foundation Trust, Public Health England (especially David Wyllie), Discovery Data Service/Endeavour Health Charitable Trust (especially David Stables), Voror Health Technologies Ltd (especially Sophie Don), NHS England (for what was NHS Digital) - for GDPR-compliant data sharing backed by individual written informed consent.

UK Biobank: UK Biobank is generously supported by its founding funders the Wellcome Trust and UK Medical Research Council, as well as the Department of Health, Scottish Government, the Northwest Regional Development Agency, British Heart Foundation and Cancer Research UK. UK Biobank genetic and phenotypic data were obtained under UK Biobank applications 43027 and 59822. Analyses of sequencing data were conducted on the Research Analysis Platform (<https://ukbiobank.dnanexus.com>). UK Biobank has ethical approval from the UK National Health Service (NHS) National Research Ethics Service (11/NW/0382).

For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

Conceptualization: MDT and CB; Methodology: KC, DJS, MDT and CB; Software: KC, RP; Validation: KC; Formal Analysis and Interpretation: KC, JG, NS, AGI, SK, EFM, RP, WH,

MM; Investigation: KC, CJ, MDT and CB; Resources: CJ, DJS, RP, HB, SRO, CE, OBVP, DAH, Genes & Health Research Team, RCF, EJH, LVW and MDT; Data Curation: KC, CB; Writing – Original Draft: KC, CJ, MDT, and CB; Writing – Review & Editing: all authors; Visualization: KC, NS; Supervision: MDT and CB; Funding Acquisition: EJH, LVW, MDT and CB.