

SUPPLEMENTARY MATERIALS

Supplementary Table 1. ICD-10 codes for pneumonia in UK Biobank	
Code	Description
B012	Varicella pneumonia
J100	Influenza with pneumonia, influenza virus identified
J111	Influenza with pneumonia, virus not identified
J120	Adenoviral pneumonia
J121	Respiratory syncytial virus pneumonia
J122	Parainfluenza virus pneumonia
J123	Human metapneumovirus pneumonia
J128	Other viral pneumonia
J129	Viral pneumonia unspecified
J13	Pneumonia due to Streptococcus pneumoniae
J14	Pneumonia due to Haemophilus influenzae
J150	Pneumonia due to Klebsiella pneumoniae
J151	Pneumonia due to Pseudomonas
J152	Pneumonia due to staphylococcus
J153	Pneumonia due to streptococcus, group B01
J154	Pneumonia due to other streptococci
J155	Pneumonia due to Escherichia coli
J156	Pneumonia due to other aerobic Gram-negative bacteria
J157	Pneumonia due to Mycoplasma pneumoniae
J158	Other bacterial pneumonia
J159	Bacterial pneumonia, unspecified
J16	Pneumonia due to other infectious organisms, not elsewhere classified
J160	Chlamydial pneumonia
J168	Pneumonia due to other specified infectious organisms
J170	Pneumonia in bacterial diseases classified elsewhere

J171	Pneumonia in viral diseases classified elsewhere
J172	Pneumonia in mycoses
J173	Pneumonia in parasitic diseases
J178	Pneumonia in other diseases classified elsewhere
J180	Bronchopneumonia, unspecified
J181	Lobar pneumonia, unspecified
J182	Hypostatic pneumonia, unspecified
J188	Other pneumonia, organism unspecified
J189	Pneumonia, unspecified
J851	Abscess of lung with pneumonia

Supplementary Table 4. Genes identified by fastBAT analysis						
Gene	CHR	Start	End	No_SNPs	Pvalue	P adj smoking phenotypes*
MAPKAP1	9	128199672	128469513	185	2.56E-05	0.0024
PBX3	9	128509616	128729655	147	2.20E-05	0.00032
ADAMTS7	15	79051544	79103773	163	1.62E-05	0.68
CHRNA3	15	78885394	78913637	84	4.00E-09	0.5
CHRNA5	15	78857861	78887611	73	3.41E-09	0.57
CHRNA4	15	78916635	78933587	83	6.94E-09	0.43
HYKK	15	78799905	78829715	65	1.68E-09	0.43
IREB2	15	78730517	78793798	97	9.80E-09	0.53
LOC646938	15	79044378	79045734	117	9.49E-06	0.43
PSMA4	15	78832746	78841563	58	4.57E-10	0.33
APOBR	16	28505969	28510291	45	1.82E-05	0.00038
CLN3	16	28477973	28503623	43	1.56E-05	0.0003
EIF3C	16	28390902	28437775	17	3.54E-06	0.000045
EIF3CL	16	28390899	28415162	14	8.09E-06	0.00011
IL27	16	28510682	28518155	51	2.59E-05	0.00052
MIR6862-1	16	28402302	28402372	6	6.76E-06	0.00011
MIR6862-2	16	28402302	28402372	6	6.76E-06	0.00011
NPIP6	16	28353838	28374181	40	2.60E-05	0.00075
Showing significant (FDR<0.05) results for fastBAST gene-based association analysis for pneumonia meta-analysis. *P values for sensitivity gene based test on the pneumonia gwas conditioned on smoking history and cigarettes per day adjusted for bonferroni corrected multiple testing (18 comparisons)						

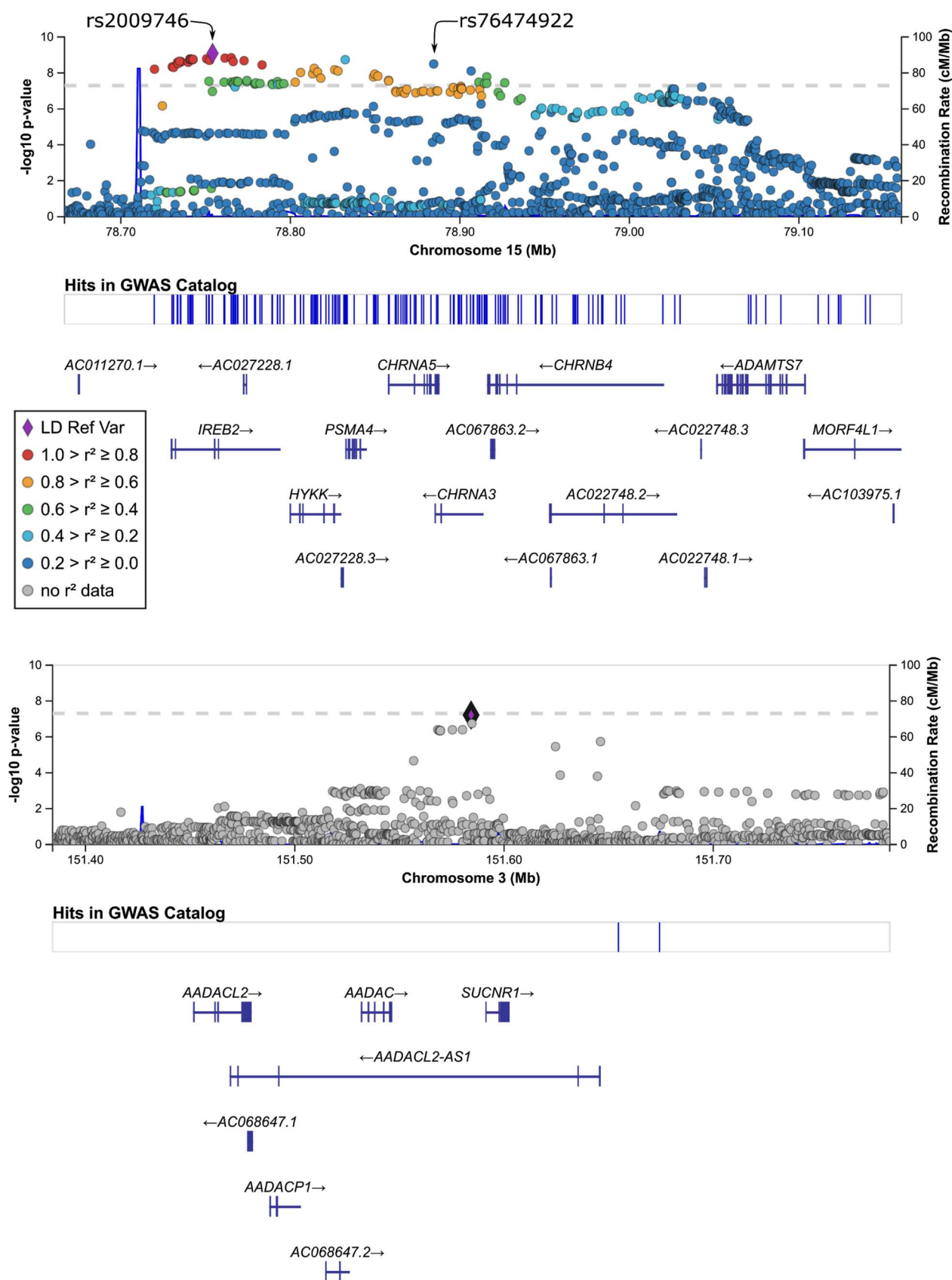
Supplementary Table 3. Colocalization of blood eQTLs with pneumonia GWASMA loci							
Gene	COLOC Posterior Probability						
	PP0	PP1	PP2	PP3	PP4	PP3+PP4	PP4/(PP3+PP4)
EIF3C	0.026	0.031	0.102	0.123	0.718	0.841	0.854
APOBR	0.431	0.523	0.016	0.020	0.010	0.029	0.325
EIF3CL	0.414	0.503	0.036	0.044	0.003	0.047	0.066
NPIPB6	0.187	0.227	0.123	0.149	0.314	0.463	0.678
IL27	0.004	0.005	0.446	0.541	0.004	0.545	0.007
CLN3	0.007	0.008	0.430	0.522	0.033	0.555	0.059
PSMA4	2.43E-09	9.75E-07	0.002	0.989	0.009	0.998	0.009
CHRNA5	0.001	0.210	0.002	0.774	0.013	0.787	0.017
IREB2	9.19E-24	3.69E-21	2.49E-03	9.97E-01	1.28E-05	0.997511	1.29E-05
HYKK	0.002	0.842	3.40E-04	0.136	0.020	0.156	0.126
PBX3	0.805	0.099	0.085	0.010	0.001	0.011	0.071
ADAMTS7	0.002	0.835	3.43E-04	0.138	0.025	0.163	0.156
MAPKAP1	0.851	0.104	0.040	0.005	3.86E-04	0.005	0.073
PP0 - no association with gene expression and pneumonia risk; PP1- association with pneumonia GWAS only; PP2 - association with gene expression only; PP3 - association with gene expression and pneumonia GWAS, but two distinct SNP; PP4 - association with gene expression and pneumonia GWAS, shared SNP; Genes not shown could not be assessed due to lack of expression in the relevant tissue or lack of eQTL data. Genes showed evidence of colocalization are in bold.							

Supplementary Table 4. Conditional analysis results for smoking associated variants	
SNP (Conditioned on)	P-value
rs16969968(rs2009746;rs588765;rs578776)	0.20477
rs588765(rs2009746;rs16969968;rs578776)	0.0698019
rs578776(rs2009746;rs588765;rs16969968)	0.588421
rs2009746(rs16969968;rs588765;rs578776)	0.00232078
Logistic regression was performed by adjusting for all relevant covariates plus the relevant SNPs shown in parenthesis. UK Biobank data only	

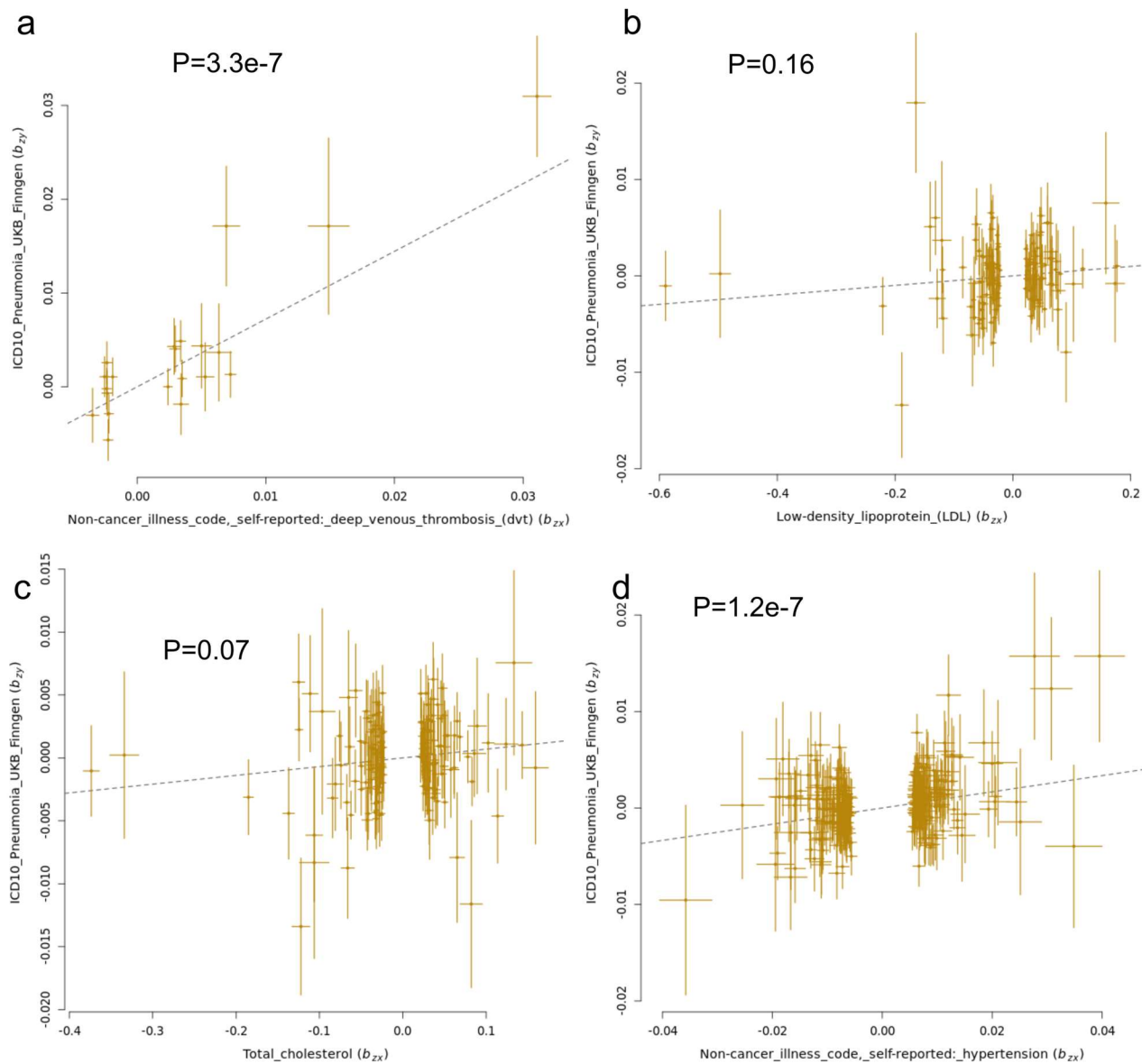
Supplementary Table 5. Results of a multivariate logistic regression for pneumonia before and after smoking adjustment

	Before adjusting			After adjusting		
	OR	std err	P	OR	std err	P
Male	0.67	0.08	2.8e-7	0.67	0.079	2.9e-7
Smoking history	NA	NA	NA	1.23	0.084	0.012
Age	1.03	0.002	1.5e-38	1.03	0.002	9.5e-31
Pneumonia PRS	1.11	0.03	0.001	1.11	0.03	0.002
Pneumonia (mtCOJO ever smoked and cigs per day) PRS	1.11	0.03	0.001	1.11	0.03	0.002

Total sample size with available data N= 5595. Std err - standard error on the logistic scale



Supplementary Figure 1. Locus Zoom plot of the genome-wide significant hits.



Supplementary Figure 2. GMR analysis of selected LCV significant traits

Pickrell plots show the results of a GMR analysis to further assess evidence for causality from the LCV results. The x-axis depicts the effect sizes for SNPs associated with an exposure of interest, the y axis shows those SNP's effect sizes on pneumonia. A significant slope (positive or negative) could be considered evidence for causality.

SUPPLEMENTARY SOFTWARE REFERENCES

1. Loh, Po-Ru, et al. "Efficient Bayesian mixed-model analysis increases association power in large cohorts." *Nature genetics* 47.3 (2015): 284.
2. Loh, Po-Ru, et al. "Mixed-model association for biobank-scale datasets." *Nature genetics* 50.7 (2018): 906-908.
3. Willer, Cristen J., Yun Li, and Gonçalo R. Abecasis. "METAL: fast and efficient meta-analysis of genomewide association scans." *Bioinformatics* 26.17 (2010): 2190-2191.
4. Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1), 76-82.
5. Yang, Jian, et al. "Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations." *Genome-wide association studies and genomic prediction*. Humana Press, Totowa, NJ, 2013. 215-236.
6. Lloyd-Jones, Luke R., et al. "Improved polygenic prediction by Bayesian multiple regression on summary statistics." *Nature communications* 10.1 (2019): 1-11.
7. Purcell, Shaun, et al. "PLINK: a tool set for whole-genome association and population-based linkage analyses." *The American journal of human genetics* 81.3 (2007): 559-575.
8. Chang, Christopher C., et al. "Second-generation PLINK: rising to the challenge of larger and richer datasets." *Gigascience* 4.1 (2015): s13742-015.
9. Zhu, Z. et al. (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* 9, 224.
10. Grotzinger, Andrew D., et al. "Genomic SEM provides insights into the multivariate genetic architecture of complex traits." *BioRxiv* (2018): 305029.