Supplementary table 1. SNPs included in the 25(OH)D allele score in the HUNT2 study (n=54580)

SNP	In gene/near gene (distance[*])	Chromosome: position	Other allele/ effect allele[†]	Effect allele[†] frequency (HUNT2)	Effect allele[†] frequency[#]	Hardy–Weinberg equilibrium *P* value	Gene function
rs2282679	*GC/*	4:71742666	G/T	0.73	0.80	0.20	vitamin D binding protein
rs12785878	*NADSYN1/DHCR7* (3.5 kb)	11:71456403	G/T	0.35	0.35	0.94	NAD synthetase 1/ 7-dehydrocholesterol reductase
rs10741657	*/CYP2R1* (1.0 kb)	11:14893332	G/A	0.39	0.31	0.36	vitamin D 25-hydroxylase

25(OH)D: 25-hydroxyvitamin D; HUNT2: The Nord-Trøndelag Health Study Survey 2; SNP: Single-Nucleotide Polymorphism

[*]SNP is located upstream of the gene

[†]Effect allele: 25(OH)D increasing allele

[#]From 1000 Genomes Phase 3 combined population

Supplementary table 2. Associations between the 25(OH)D allele score and potential confounders in the HUNT2 study (n=54580)

Baseline variables[*]	Coefficient[†]	95% CI	*P* value
Age (difference in years)	0.06	(-0.06 to 0.18)	0.30
Sex (men vs. women)	-0.004	(-0.02 to 0.01)	0.57
Active smoking (ever vs. never)	-0.004	(-0.02 to 0.01)	0.61
Pack-years of active smoking (difference)	-0.01	(-0.03 to 0.01)	0.19
Family history of cancer (yes vs. no)	0.004	(-0.01 to 0.02)	0.66
Education (years) (≥10 vs. <10)	-0.001	(-0.02 to 0.02)	0.93
Economic difficulties (yes vs. no)	-0.01	(-0.03 to 0.01)	0.41
Body mass index (BMI, kg/m^2) (≥25 vs. <25)	0.01	(-0.00 to 0.03)	0.09
Physical activity (active vs. inactive)	0.0002	(-0.02 to 0.02)	0.98
Alcohol consumption (times/month) (≥1 vs. never)	-0.02	(-0.03 to -0.00)	0.04
Chronic bronchitis (yes vs. no)	-0.01	(-0.05 to 0.03)	0.77

25(OH)D: 25-hydroxyvitamin D; CI: confidence interval; HUNT2: The Nord-Trøndelag Health Study Survey 2

[*]Data of baseline variables were collected by questionnaires or clinical examination in HUNT2. The classification of each covariate was described in detail in a previous study [1].

[†]Coefficient was derived from linear regression for continuous variables and from logistic regression for categorical variables.

Supplementary table 3. MR estimates of a 10% increase in genetically determined 25(OH)D and risk of lung cancer overall and histologic types in a two-sample MR analysis*

Outcome	IVW method						Weighted median method		
	MR estimate HR (95% CI)		*P* value	I^2 (95% CI)		*P* value of Q statistic	MR estimate HR (95% CI)		*P* value
Lung cancer overall	0.99	(0.88 to 1.12)	0.85	0.00	(0.00 to 0.20)	0.88	0.99	(0.87 to 1.13)	0.91
SCLC	0.87	(0.63 to 1.20)	0.39	0.41	(0.00 to 0.82)	0.18	0.90	(0.63 to 1.29)	0.56
Adenocarcinoma	0.88	(0.70 to 1.10)	0.27	0.00	(0.00 to 0.68)	0.72	0.89	(0.71 to 1.13)	0.35
Squamous cell carcinoma	0.89	(0.69 to 1.15)	0.37	0.00	(0.00 to 0.89)	0.39	0.89	(0.68 to 1.17)	0.40
Other/unknown subtypes	1.20	(0.98 to 1.47)	0.07	0.00	(0.00 to 0.24)	0.87	1.22	(0.99 to 1.52)	0.06

25(OH)D: 25-hydroxyvitamin D; CI: confidence interval; HR: hazard ratio; IVW: inverse-variance weighted; MR: Mendelian randomization; SCLC: small cell lung cancer

*Summarized data of SNPs–25(OH)D association was derived from the study of Vimaleswaran *et al* [2] (n≈35000); summarized data of SNPs–outcome association was derived from the HUNT2 study (n=54580)

Supplementary table 4. MR-Egger pleiotropy test of associations between a 10% increase in genetically determined 25(OH)D and risk of lung cancer overall and histologic types in a two-sample MR analysis*

Outcome	MR-Egger method		
	Intercept (95% CI)		*P* value
Lung cancer overall	-0.01	(-0.16 to 0.14)	0.90
SCLC	-0.08	(-0.83 to 0.66)	0.83
Adenocarcinoma	-0.10	(-0.39 to 0.19)	0.49
Squamous cell carcinoma	0.22	(-0.11 to 0.55)	0.19
Other/unknown subtypes	-0.05	(-0.31 to 0.20)	0.68

25(OH)D: 25-hydroxyvitamin D; CI: confidence interval; MR: Mendelian randomization; SCLC: small cell lung cancer

*Summarized data of SNPs–25(OH)D association was derived from the study of Vimaleswaran *et al* [2] (n≈35000); summarized data of SNPs–outcome association was derived from the HUNT2 study (n=54580)

Supplementary table 5. Mendelian randomization estimates of the associations between one pack-year increase of genetically determined smoking and risk of lung cancer overall and histologic types among ever smokers in the HUNT2 study (n=26815)

	Number of cases	MR estimate HR*	95% CI	*P* value
Lung cancer overall	548	1.48	(1.29 to 1.70)	3.48×10^{-8}
SCLC	80	1.34	(0.91 to 1.98)	0.14
Adenocarcinoma	149	1.87	(1.44 to 2.44)	3.41×10^{-6}
Squamous cell carcinoma	115	1.06	(0.78 to 1.45)	0.70
Other/unknown subtypes	204	1.55	(1.24 to 1.94)	9.61×10^{-5}

CI: confidence interval; HR: hazard ratio; HUNT2: The Nord-Trøndelag Health Study Survey 2; MR: Mendelian randomization; SCLC: small cell lung cancer

*Two-stage method was applied [3] using rs1051730 as instrumental variable for smoking quantity [4]. Firstly, based on the linear regression of per effect allele (A) of rs1051730 on pack-years of smoking, the predicted values of pack-years of smoking were generated. Secondly, Cox regression analyses of predicted pack-year values as exposure on incidence of lung cancer overall and histologic types were performed. Robust standard errors were applied for both regressions.

References

1. Sun YQ, Langhammer A, Wu C, Skorpen F, Chen Y, Nilsen TIL, Romundstad PR, Mai XM. Associations of serum 25-hydroxyvitamin D level with incidence of lung cancer and histologic types in Norwegian adults: a case-cohort analysis of the HUNT study. *Eur J Epidemiol* 2017.
2. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, Cooper JD, Dastani Z, Li R, Houston DK, Wood AR, Michaelsson K, Vandenput L, Zgaga L, Yerges-Armstrong LM, McCarthy MI, Dupuis J, Kaakinen M, Kleber ME, Jameson K, Arden N, Raitakari O, Viikari J, Lohman KK, Ferrucci L, Melhus H, Ingelsson E, Byberg L, Lind L, Lorentzon M, Salomaa V, Campbell H, Dunlop M, Mitchell BD, Herzig KH, Pouta A, Hartikainen AL, Streeten EA, Theodoratou E, Jula A, Wareham NJ, Ohlsson C, Frayling TM, Kritchevsky SB, Spector TD, Richards JB, Lehtimaki T, Ouwehand WH, Kraft P, Cooper C, Marz W, Power C, Loos RJ, Wang TJ, Jarvelin MR, Whittaker JC, Hingorani AD, Hypponen E. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. *PLoS Med* 2013: 10(2): e1001383.
3. Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, Meng S, Granell R, Smith GD, Didelez V. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *Am J Epidemiol* 2011: 173(12): 1392-1403.
4. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010: 42(5): 441-447.