**Supplementary Methods**

**Exposure and Outcome GWAS Description**

**-Inflammatory Protein GWAS**

This study consisted of up to 30,931 individuals from 14 studies with measurements on 90 proteins measured using the Cardiovascular-I panel on the Olink platform as part of the SCALLOP consortium (Average protein GWAS sample size: 17,747). GWAS analyses were conducted for inverse-rank normalised/standardised protein concentrations using an additive model in a linear regression with adjustment for population structure using principal components of ancestry in each SCALLOP study. All study results were combined using a fixed effects meta-analysis with inverse-variance weighting. Further details are available elsewhere1.

**-Lung cancer GWAS**

The study involved 85,716 participants (29,266 cases; 56,450 controls) from several genetic studies managed by the Transdisciplinary Research of Cancer in Lung of the International Lung Cancer Consortium (TRICL-ILCCO) and the Lung Cancer Cohort Consortium (LC3). GWAS analysis involved logistic regression models for each SNP (imputation was based on the 1000 Genomes Project Phase 3 panel) that included the 3 first principal components from a principal component analysis to infer genetic ancestry by genotype. Fixed effects meta-analysis with inverse-variance weighting was used to integrate results from the participating studies; description of all studies that took part of the meta-analysis, their inclusion/exclusion criteria, genotype quality control and detailed analytical plan are described elsewhere.2

**Protein-Tumor Immune Analyses**

**- Quality Control for Samples in the Cancer Genome Atlas**

The Cancer Genome Atlas14 is a joint effort between the National Cancer Institute and the National Human Genome Research Institute that characterised over 10,000 primary cancers across 33 cancer types. Patient samples were filtered to exclude the following TCGA3 non-rescinded annotations: "Item flagged DNU", "Administrative Compliance", "Item does/may not meet study protocol", "Qualified in error", "BCR Notification", "Normal tissue origin incorrect", "Subject withdrew consent", "Normal class but appears diseased", "Duplicate item", "Tumor tissue origin incorrect", "Tumor type incorrect", "Genotype mismatch", "Permanently missing item or object" and "Qualification metrics changed".

**Protein PRS and Lung Cancer Risk**

**-Study Samples and PRS method**

PRS analyses were conducted in the UK Biobank, a large prospective study started in 2006 with genotype information in over 500,000 participants age between 40 and 60 at recruitment. For this study, sample selection and SNP filters were non-white British ancestry, sex discrepancy, aneuploidy, and 2nd degree relatives leaving 344,364 participants for analyses. Genetically predicted protein concentrations for proteins with significant MR associations were generated using the LDPred algorithm4 in Folkersen et al1 and estimated in UK Biobank using PRSice5.

**Pre-diagnostic Protein Concentrations and Lung Cancer Risk**

**-Study Samples**

Measurements for up to 1,182 proteins were available for a nested case-control study as part of an ongoing lung cancer protein project (U19 CA203654) using the Olink platform within six prospective cohorts. Pre-diagnostic measurements for proteins with promising MR results were extracted from these data to facilitate triangulation for lung cancer aetiology.

*EPIC*

The European Investigation into Cancer and Nutrition (EPIC) study is a large multi-centre prospective cohort that recruited participants between 1992 and 1998 and gather both questionnaire data and blood samples at recruitment from participants. For the case-control study, participants were taken from the 238,816 individuals from the centres in Netherlands, UK, France, Germany, Spain, and Italy who donated a blood sample at study recruitment. Further information regarding EPIC is provided elsewhere6.

*NSHDS*

The Northern Sweden Health and Disease Study (NSHDS)7 includes several prospective cohorts. The current study included study participants from the Västerbotten Intervention Project (VIP), which is a sub-cohort within NSHDS. In 1985, when VIP was started, all residents in the Västerbotten County were invited to participate by attending a health check-up at 40, 50 and 60 years of age. Participants completed a self-administered questionnaire and donated blood samples (fasting) for future research purposes.

*HUNT*

The Trøndelag Health Study (HUNT)8 is a large population-based cohort started in the 1980’s, covering 125,000 Norwegian participants from three sub-cohorts: HUNT1 (1984-86), HUNT2 (1995-97) and HUNT3 (2006-08). The study was primarily set up to address arterial hypertension, diabetes, screening of tuberculosis, and quality of life, however, this has expanded to include additional diseases.  The HUNT Study includes data from questionnaires, interviews, clinical measurements and biological samples (blood and urine).

*CPS-II*

The Cancer Prevention Study II (CPS-II)9 is a prospective mortality study of approximately 1.2 million American men and women recruited in all 50 states, the District of Columbia, and Puerto Rico beginning in 1982. Participants completed a questionnaire for lifestyle and anthropometric characteristics and 40,000 donated a blood sample at recruitment. During the 24 years of completed mortality follow-up currently available for this cohort (1982-2006), 491,188 deaths have occurred; cause of death has been obtained for 99.3% of all deaths.

*SCHS*

The Singapore Chinese Health Study (SCHS)10 is a population-based prospective cohort study of 63,000 Chinese men and women living in Singapore recruited between 1993 and 1998. The primary goal of this longitudinal cohort study is to elucidate the role of diet and its interaction with genetic and genomic factors in the causation of cancer and other chronic diseases. Each participant completed a lifestyle and anthropometry questionnaire and donated a blood and urine.

*MCCS*

The Melbourne Collaborative Cohort Study (MCCS)11 is a prospective population cohort of 41,513 residents of the Melbourne metropolitan area recruited between 1990 and 1994 with the aim of investigating the roles of diet and lifestyle in causing cancer and other non-communicable diseases. Participant completed a lifestyle and anthropometry questionnaire and a majority of participants also donated a blood sample.

- **Matching Criteria**

Controls were individually matched to the index case, cohort, study centre (where relevant), sex, date of blood collection (±1 month, relaxed to ±3 months for sets without available controls), and date of birth (±1 year, relaxed to ±3 years), and smoking status in four categories: former smokers with <10 or ≥10 years since quitting, and current smokers with <15 or ≥15 cigarettes smoked per day. Quartiles were fourths delimited in the controls.

- **Protein Measurements**

The abundance of proteins was measured in plasma or serum samples by multiplex proximity extension assay by Olink Proteomics AB (Uppsala, Sweden) using a Fluidigm Biomark reader (Fluidigm Corporation, USA). Paired case-control samples were randomly allocated over 96-well plates. Protein levels were measured as normalized protein expression (Ct values with corrections for assay variation) on the log2 scale. We excluded samples that did not pass laboratory quality control criteria (100% of values were missing, N = 3). As per advice, samples below the lower limit of detection (LOD) were replaced with LOD/sqrt(2)12. Additionally, we excluded samples that did not pass quality control criteria provided by the manufacturer.

1. Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* (2020). doi:10.1038/s42255-020-00287-2

2. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* (2017). doi:10.1038/ng.3892

3. Weinstein, J. N. *et al.* The cancer genome atlas pan-cancer analysis project. *Nature Genetics* (2013). doi:10.1038/ng.2764

4. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* (2015). doi:10.1016/j.ajhg.2015.09.001

5. Choi, S. W., Mak, T. S. H. & O’Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols* (2020). doi:10.1038/s41596-020-0353-1

6. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* (2002). doi:10.1079/phn2002394

7. Hallmans, G. *et al.* Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort- evaluation of risk factors and their interactions. *Scand. J. Public Health* (2003). doi:10.1080/14034950310001432

8. Krokstad, S. *et al.* Cohort profile: The HUNT study, Norway. *Int. J. Epidemiol.* (2013). doi:10.1093/ije/dys095

9. Calle, E. E. *et al.* The american cancer society cancer prevention study II nutrition cohort: Rationale, study design, and baseline characteristics. *Cancer* (2002). doi:10.1002/cncr.10197

10. Hankin, J. H. *et al.* Singapore Chinese Health Study: Development, Validation, and Calibration of the Quantitative Food Frequency Questionnaire. *Nutr. Cancer* (2001). doi:10.1207/S15327914nc392\_5

11. Milne, R. L. *et al.* Cohort Profile: The Melbourne Collaborative Cohort Study (Health 2020). *Int. J. Epidemiol.* (2017). doi:10.1093/ije/dyx085

12. Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* (2014). doi:10.1371/journal.pone.0095192