



## Research article

## Serum protein profiling of lung, pancreatic, and colorectal cancers reveals alcohol consumption-mediated disruptions in early-stage cancer detection

Hanjun Lee<sup>a,b,c,d</sup>, Seo Yihl Kim<sup>a</sup>, Dong Wook Kim<sup>a</sup>, Young Soo Park<sup>e,f</sup>, Jin-Hyeok Hwang<sup>e,f,\*\*\*</sup>, Sukki Cho<sup>g,h,\*\*</sup>, Je-Yoel Cho<sup>a,i,\*</sup><sup>a</sup> Department of Biochemistry, College of Veterinary Medicine, BK 21 PLUS, Seoul National University, Seoul 08826, South Korea<sup>b</sup> Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, United States<sup>c</sup> Broad Institute of MIT and Harvard, Cambridge, MA 02139, United States<sup>d</sup> Massachusetts General Hospital Cancer Center, Charlestown, MA 02129, United States<sup>e</sup> Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam 13620, South Korea<sup>f</sup> Department of Internal Medicine, Seoul National University College of Medicine, Seoul 03080, South Korea<sup>g</sup> Department of Thoracic and Cardiovascular Surgery, Seoul National University Bundang Hospital, Seongnam 13620, South Korea<sup>h</sup> Department of Thoracic and Cardiovascular Surgery, Seoul National University College of Medicine, Seoul 03080, South Korea<sup>i</sup> Comparative Medicine Research Center, Seoul National University, Seoul 08826, South Korea

## ARTICLE INFO

## Keywords:

Cancer  
Lung cancer  
Pancreatic cancer  
Colorectal cancer  
Alcohol  
Biomarker  
Serum protein

## ABSTRACT

While the link between serum proteins and cancer has been studied in an effort to enable early-stage cancer detection, factors that might perturb this link has been poorly understood. To ask this question, we performed serum protein profiling on a prospective cohort of 601 individuals with or without lung, pancreatic, or colorectal cancers and identified ten distinct serum protein signatures with distinct link to the patient metadata. Importantly, we discovered that a positive history of alcohol consumption is a major factor that diminishes the sensitivity of serum protein-mediated liquid biopsy in early-stage malignancies, resulting in a 44% decline in the sensitivity of detecting American Joint Committee on Cancer (AJCC) stage I malignancies. Our data provide evidence that patient lifestyle can affect the sensitivity of liquid biopsy and suggest the potential need for abstinence from alcohol before measurement during serum protein-based cancer screening.

## 1. Introduction

The detection of cancer in its early stage is considered one of the holy grails of patient care in the fight against the disorder. One of the reasons behind this popular notion is because earlier stages of malignancies generally arise from a few driver mutations and are comprised of a less heterogeneous population of cancer cells, which leads to a better prognosis of patients after treatment [1]. Previous studies have thus investigated changes that specifically occur at the earliest stages of neoplastic progression and identified inflammation [2, 3, 4] and angiogenesis [5, 6] as notable examples. Inflammation is considered to contribute to neoplastic progression via creating a tumor-promoting microenvironment [4, 6], which often entails a constitutively active angiogenic switch in the lesion. This change typically commences during the premalignant

stage of cancer [5], and is often probed to study early-stage cancer detection.

At the intersection between angiogenesis and inflammation lies serum proteins comprised of diverse growth factors, cytokines, and soluble forms of surface receptors. Hence, serum proteins have long been studied as potential diagnostic tools for cancer [7, 8, 9, 10], and increasing number of serum protein-based liquid biopsies have entered large-scale clinical trials [11, 12, 13]. Despite an extensive effort to establish a link between serum proteins and the likelihood of neoplastic progression in individuals, factors that can affect or even disrupt this link have been poorly understood.

Herein, we performed an extensive serum protein profiling on a prospective cohort that involved 601 individuals with or without lung, pancreatic, or colorectal cancers and identified positive history of alcohol

\* Corresponding author.

\*\* Corresponding author.

\*\*\* Corresponding author.

E-mail addresses: [woltoong@snu.ac.kr](mailto:woltoong@snu.ac.kr) (J.-H. Hwang), [tubincho@snu.ac.kr](mailto:tubincho@snu.ac.kr) (S. Cho), [jeych@snu.ac.kr](mailto:jeych@snu.ac.kr) (J.-Y. Cho).<https://doi.org/10.1016/j.heliyon.2022.e12359>

Received 13 June 2022; Received in revised form 20 October 2022; Accepted 7 December 2022

2405-8440/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

consumption as a key factor diminishing the sensitivity of serum protein-based blood tests in early-stage cancers. The perturbation was driven by a positive correlation between serum protein concentrations of cancer patients who have consumed alcohol in the past and those of healthy individuals without cancer, which led to the reversing of the alterations induced by cancer occurrence alone. In contrast, the consideration of the individual's history of alcohol consumption enhanced the sensitivity of these blood tests, providing the evidence that patient lifestyle can strongly affect cancer detection by liquid biopsy. These data suggest the potential need for abstinence from alcohol in serum protein-based cancer screening.

## 2. Results

### 2.1. Identification of serum protein signatures in cancer patients

Our study involved 601 individuals with or without three of the most lethal types of cancers that account for 39% of all cancer mortalities: lung, pancreas, and colorectum [14]. Fifteen out of 420 cancer patients were later determined as having a non-primary tumor, of which protein concentration profiles we excluded for further investigations (Table S1). Serum protein markers previously described to be associated with at least one of the three cancer types studied were identified based on an exhaustive literature search (Table S2). Concentrations of these forty-eight protein biomarkers were determined by a multiplex immunoassay [15], which resulted in a total of 28,848 measurements. Clinical metadata of participants, which included sex, age, body mass index (BMI), history of tobacco smoking, history of alcohol consumption, prior history of malignancy, family history of malignancy, diabetes, and hypertension, were prospectively collected for analysis. Raw data are presented in Table S3.

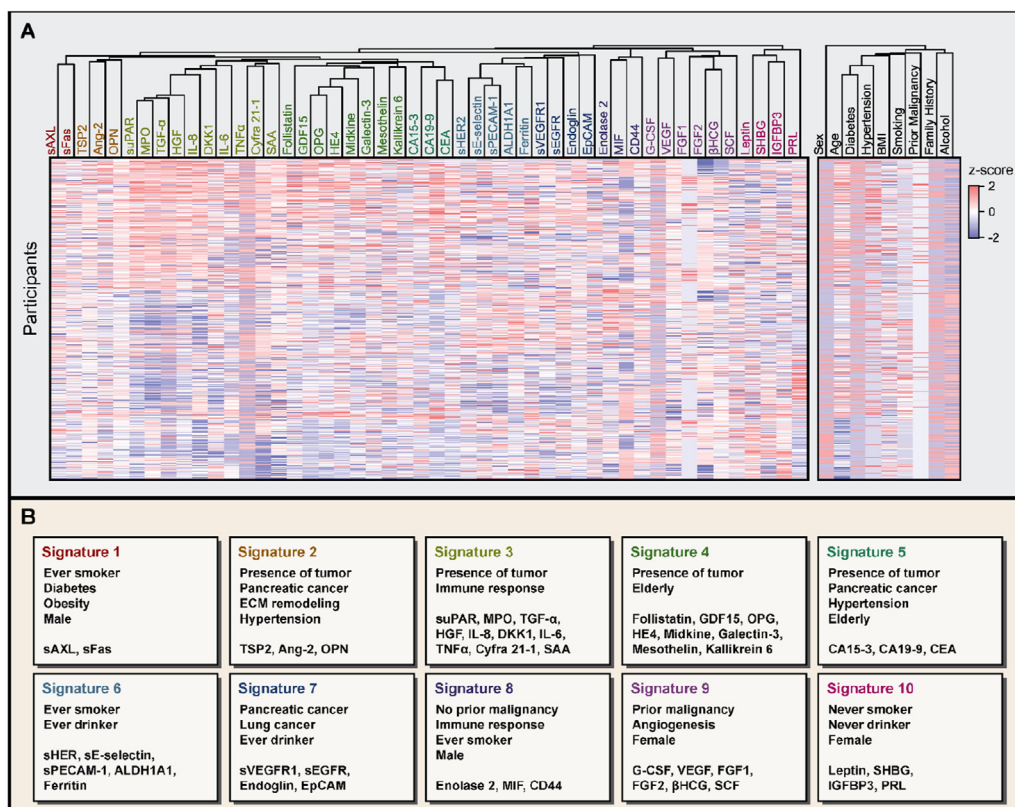
To establish serum protein signatures in cancer patients, we performed cohort-wide hierarchical clustering of serum proteins (Materials and methods). Euclidean distance-based hierarchical clustering revealed ten prominent serum protein signatures (signatures 1–10; Figure 1A).

Each signature had varying number of member proteins, which ranged from the minimum of two to the maximum of ten. Complete list of member proteins for each signature and its signature-specific attributes are described in Figure 1B.

Signatures covered expansive aspects of cancer biology. Signatures 3 and 8 were strongly associated with immune response, with cytokines being the predominant member of these signatures. IL-6, IL-8, and TNF $\alpha$ , which are well-known proinflammatory cytokines that are often co-regulated [16], all belonged to signature 3. Analogously, both MIF and its co-receptor CD44, which are often overexpressed under inflammatory stimuli [17], belonged to signature 8. Signature 9 was highly enriched with growth factors associated with angiogenesis, including VEGF and FGFs [18], while signature 2 was enriched with serum proteins associated with extracellular matrix (ECM) remodeling, including angiopoietin-2 (Ang-2), osteopontin (OPN), and thrombospondin 2 (TSP2) [19, 20, 21]. Overall, these links to diverse cancer-related processes corroborate the biological relevance of classifying co-regulated serum biomarkers into a set of serum protein signatures.

### 2.2. Patient metadata strongly regulates serum protein signatures

To further understand the biological relevance of these signatures, we performed metadata analysis on nine clinical parameters prospectively collected (Materials and methods). For each parameter, binarization was performed to enable statistical comparison between two distinct groups of individuals with different clinical characteristics: female versus male (sex), non-elderly versus elderly (age), non-diabetic versus diabetic (diabetes), normotensive versus hypertensive (hypertension), lean versus obese (BMI), never smoker versus ever smoker (tobacco smoking), no prior malignancy versus prior malignancy (prior history of malignancy), no family history versus family history (family history of malignancy), and never drinker versus ever drinker (alcohol consumption). Criteria for the binarization are further detailed in Materials and methods. In accordance with previous literatures [22, 23, 24, 25, 26, 27, 28], sex, tobacco smoking, and hypertension significantly contributed to



**Figure 1.** Multicancer cohort reveals ten prominent serum protein signatures. (A) Heatmap of serum protein concentrations and individual metadata in the cohort. Each column has been hierarchically clustered using the Euclidean distance. Ten co-regulated signatures of serum proteins are color-coded. Each row represents 586 participants who met the inclusion criteria. (B) Signatures and their signature-specific attributes. Members of each signature are presented in panels. SeroHet signatures encompassed expansive aspects of cancer biology, including stage, cancer type, immune response, angiogenesis, and extracellular matrix remodeling.

interpatient serum protein heterogeneity in participants, as shown by the signature-wise fold changes (Figure 2A) and their corresponding  $p$ -values (Figure 2B). By contrast, family history had the least contribution to interpatient serum protein heterogeneity (Figure S1), with signature 6 being the only signature that was significantly associated ( $p$ -value = 0.0108). Signature 6 was associated with all metadata studied except for the age of participants and was strongly overrepresented in ever-smokers ( $p$ -value < 0.001), providing a potential explanation for the observed negative correlation between signature 6 expression and family history of malignancy (Figure 2A). Individuals without the family history of malignancy may require stronger association with potent carcinogenic factors, such as tobacco smoking, to confer carcinogenesis. For individuals without the family history of malignancy, tobacco smoking was a stronger predictor of a potential neoplastic progression compared to age.

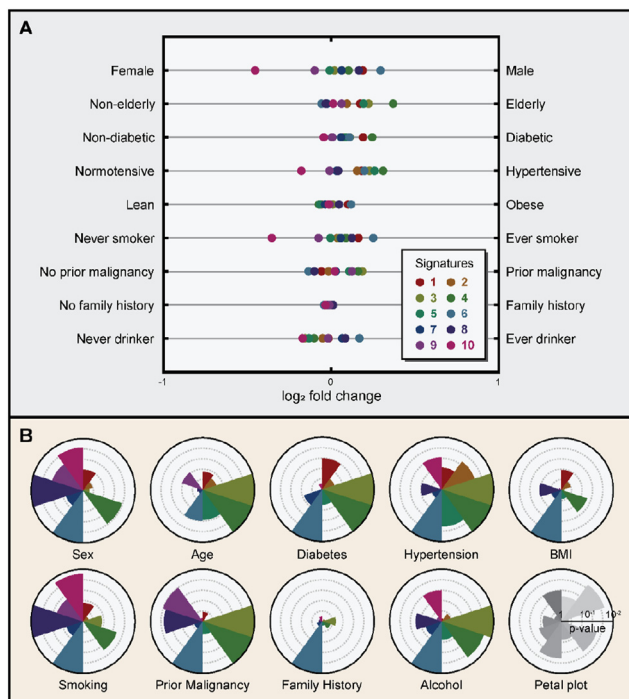
Many previously reported associations between serum proteins and patient metadata have been reproduced, substantiating the validity of our observations. For instance, serum protein markers previously known to correlate with tobacco smoking, including aldehyde dehydrogenase 1 family member A1 (ALDH1A1) [26, 28] and ferritin [27], were rediscovered by the signature analysis in this cohort. Furthermore, lung cancer patients exhibited 50% increased likelihood of positive history of tobacco smoking compared to pancreatic or colorectal cancer patients ( $p$ -value < 0.001, Fisher's exact test). In general, never-smokers overexpressed signature 10, while ever-smokers overexpressed signatures 6 and 8. As previously described, sex also had a strong influence on serum protein concentrations, with female participants being enriched with signatures 9 and 10, while male participants were enriched with signatures 1, 6, and 8. Prolactin (PRL) and sex hormone binding globulin (SHBG), which are two well-known sex hormone-associated serum proteins typically overexpressed in female individuals [22, 25], could also be rediscovered in our cohort. Alcohol consumption also had a substantial

influence on the serum protein signatures of individuals, with signature 6 being overrepresented in ever-drinkers and signatures 3, 4, and 10 being overrepresented in never-drinkers. The observed positive correlation between alcohol consumption and serum ferritin level corroborates previous reports [29, 30]. ALDH1A1, an aldehyde dehydrogenase known to function in ethanol detoxification [31, 32], also positively correlated with chronic alcohol consumption, corroborating our analysis.

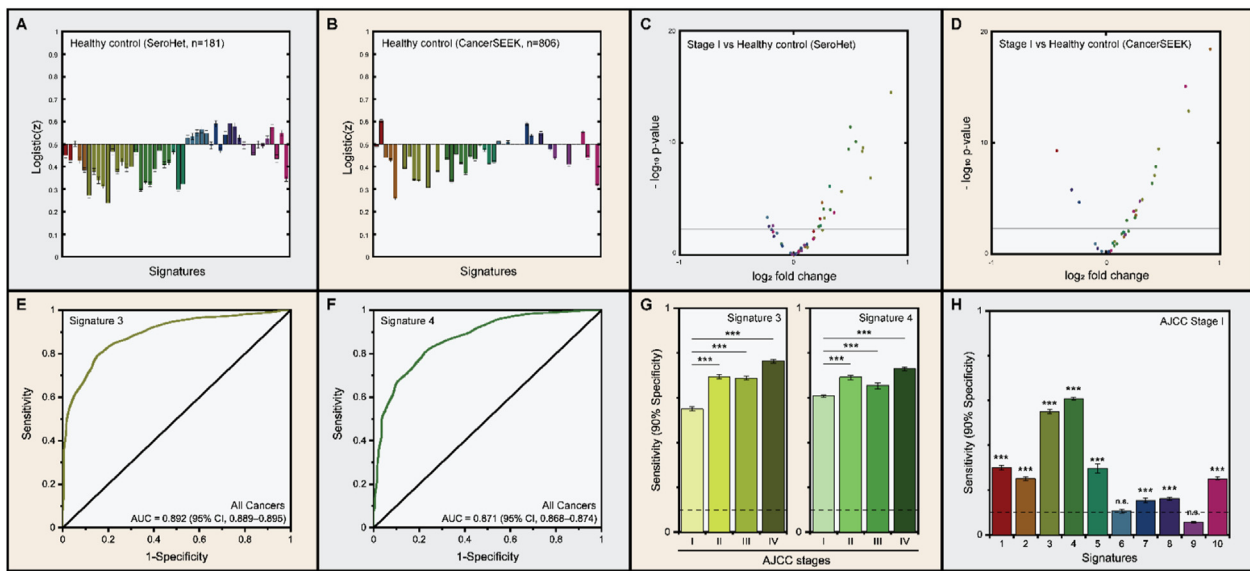
### 2.3. Serum protein signatures 2–5 are diagnostic of early-stage cancers

To understand the role of each signature in serum protein-based cancer detection, we compared signatures of the American Joint Committee on Cancer (AJCC) stage I cancer patients to those of healthy individuals without cancer. In order to confirm that the observed serum protein signatures are independent of cohorts, we also analyzed signatures in the CancerSEEK cohort [11], which also included patients with lung, pancreatic, and colorectal cancers. Strong positive correlation was observed between serum protein signatures from individuals with lung, pancreatic, or colorectal cancers in the two cohorts (Spearman's  $\rho$  = 0.716,  $p$ -value < 0.001; Figures S2 and 3), indicating that the observed differences in serum protein signatures are reproducible. In both cohorts, signatures 2–5 were overexpressed in cancer patients compared to healthy individuals without cancer (Figure 3A, B). Considering that signature 2 is associated with extracellular matrix remodeling and that signature 3 is associated with inflammatory immune response, the observed serum protein difference between cancer patients and healthy individuals are highly predicted. Volcano plot between healthy individuals and clinical stage I cancer patients in both cohorts revealed more serum proteins upregulated by the presence of tumor compared to vice versa (Figure 3C, D), suggesting the potential role of these serum protein signatures in early-stage cancer detection. Indeed, all serum protein included in the CancerSEEK detection algorithm, except for PRL, belonged to signatures 2–5 [11]. Strong modulation of serum level of PRL by sex and tobacco smoking [22, 23, 24] prevented its membership to these signatures, despite its consistent overexpression in cancer patients from both cohorts. Since PRL has the least similarity to other members of signature 10, as shown by the Euclidean distance between each serum protein (Figure 1A), further expansion of serum biomarker panels may reveal novel cancer-associated serum protein signature containing PRL.

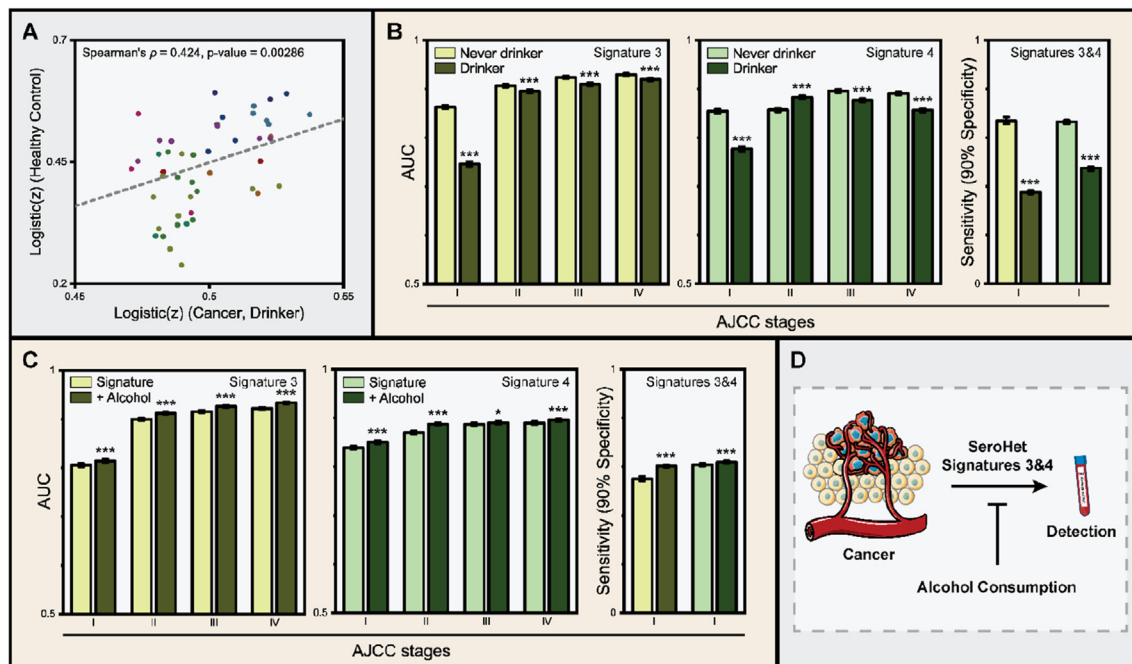
Assessment of the potential of signatures in early-stage cancer detection was done by a logistic regression (Materials and methods), as has been previously described [11, 12]. To ensure the reproducibility of our observations, we performed 1,000 iterations of 10-fold cross-validation for each signature (Figure S4). Notably, signatures 3 and 4, which had the greatest number of member proteins, were the most sensitive in detecting cancer, regardless of the cancer type (Figure S5). The average area under the receiver operating characteristic (ROC) curve, hereafter abbreviated as AUC, was 0.892 (95% confidence interval (CI), 0.889–0.895; Figure 3E) for signature 3 and 0.871 (95% CI, 0.868–0.874; Figure 3F) for signature 4. The detection performance increased with cancer progression in both serum protein signatures (Figure 3G), which is a phenomenon that has been consistently reported in serum protein-based liquid biopsies [11, 12, 13, 33, 34]. The average sensitivity of clinical stage I cancer detection at 90% specificity was 55% (95% CI, 54–56%) for signature 3 and 61% (95% CI, 60–61%) for signature 4, which were highest among all serum protein signatures studied (Figure 3H). Metadata previously known to positively correlate with cancer occurrence, which included the history of prior malignancy, aging [14], hypertension [35, 36, 37], and diabetes [38], were also enriched with signatures 3 and 4. By contrast, signatures 6 and 9 exhibited poor performance in detecting cancer, suggesting that angiogenic growth factors and soluble forms of cell adhesion molecules are relatively poor diagnostic markers for cancer. The observed poor association between signature 6 and the presence of tumor was despite signature 6 being the most upregulated signature in ever-smokers (Figure 2A).



**Figure 2.** Association of signatures to patient metadata. (A) Forest plot showing the  $\log_2$  fold changes for each signature. Signatures are color-coded and represented as a single dot. Each row represents different types of patient metadata studied. (B) Petal plot showing the  $-\log_{10}$   $p$ -values for each signature. The scale of the axis is indicated in the bottom right. Each signature is presented in a clockwise order, beginning from the top. Signatures are color-coded. The outermost circle represents the threshold for the statistical significance ( $p$ -value < 0.01).



**Figure 3.** Early-stage cancer detection by signatures. (A) Logistic(z) values of healthy individuals compared to cancer patients in the current study. (B) Logistic(z) values of healthy individuals compared to cancer patients in CancerSEEK cohort. Signatures 2–5 were underrepresented in healthy individuals compared to cancer patients in both cohorts. (C) Volcano plot between clinical stage I cancer patients and healthy individuals without cancer in SeroHet cohort. The horizontal axis indicates the  $\log_2$  fold change in the normalized protein concentration of AJCC stage I cancers compared to healthy individuals. (D) Volcano plot between clinical stage I cancer patients and healthy individuals without cancer in CancerSEEK cohort. The horizontal axis indicates the  $\log_2$  fold change in the normalized protein concentration of AJCC stage I cancers compared to healthy individuals. Horizontal line represents the Bonferroni-corrected threshold for p-value at a target family-wise error rate of 0.05. Upregulation of signatures 3 and 4 in clinical stage I cancer patients are evident in both cohorts. (E) ROC curves on the detection of cancer using SeroHet signature 3. AUC value of the curve was 0.892 (95% CI, 0.889–0.895). (F) ROC curves on the detection of cancer using signature 4. AUC value of the curve was 0.871 (95% CI, 0.868–0.874). (G) Average sensitivity at 90% specificity of detecting cancer with each AJCC stage using signatures 3 and 4. The sensitivity of the classifier positively correlated with cancer progression in both cases. (H) Average sensitivity at 90% specificity of detecting AJCC stage I cancer using each signature. SeroHet signatures 3 and 4 were highly diagnostic of cancer at AJCC stage I. Error bar, 95% CI.



**Figure 4.** Alcohol consumption perturbs cancer detection by liquid biopsy. (A) Correlation between serum protein profiles of healthy individuals without cancer and cancer patients with the history of chronic alcohol intake. Spearman's  $\rho$  was 0.424 (p-value = 0.00286), indicating that ever-drinker with cancer phenocopy healthy individuals without cancer. (B) Performance of cancer detection using signatures 3 and 4 by the history of alcohol consumption. Individuals with positive history of chronic alcohol intake underperformed in these blood tests. (C) Performance of cancer detection with and without the binary parameter on the history of chronic alcohol consumption. The inclusion of the parameter enhanced the performance of blood tests in all stages. (D) Schematic diagram illustrating the effect of chronic alcohol consumption in liquid biopsy-based cancer detection. Error bar, 95% CI.

#### 2.4. Chronic alcohol consumption perturbs the performance of liquid biopsy

Strikingly, participants with the history of alcohol consumption showed extensive downregulation of signatures 3 and 4 (Figure 2A), which may have detrimental effects on liquid biopsy-based cancer detection. To assess whether alcohol consumption perturbs cancer detection, we first performed correlation analysis of serum protein concentrations of cancer patients with the history of chronic alcohol consumption and healthy individuals without cancer. Indeed, we observed a positive correlation between the two groups of participants (Spearman's  $\rho = 0.424$ ,  $p$ -value = 0.00286), indicating that cancer patients who consistently consumed alcohol phenocopy healthy individuals, at least in these cancer-associated serum biomarkers studied. Intrigued by this result, we re-analyzed ROC curves for signatures 3 and 4 and dissected them based on the patient's history of alcohol consumption. To our surprise, consistent under-performance of cancer detection was observed in cancer patients with the history of alcohol consumption (Figure 4B). Alcohol consumption had the strongest detrimental effect in early-stage cancer detection, as the AUC for signatures 3 and 4 were decreased by 14% and 9% in clinical stage I cancer patients, respectively, upon chronic alcohol consumption. In terms of the average sensitivity at 90% specificity, signature 3 exhibited a striking 44% decline ( $p$ -value < 0.001), while signature 4 exhibited 29% decline upon chronic alcohol uptake ( $p$ -value < 0.001).

Based on this observation, we wondered whether the inclusion of the history of alcohol consumption as a novel binary parameter could enhance the performance of cancer detection by liquid biopsy. To test this hypothesis, we performed 1,000 iterations of 10-fold cross-validation in the presence of the parameter and compared its performances to those without (Materials and methods). The performance of liquid biopsies was significantly enhanced with the inclusion of the new parameter (Figure 4C), regardless of the clinical stage of the malignancy. The average sensitivity at 90% specificity was enhanced by 10% in signature 3 ( $p$ -value < 0.001) and by 2% in signature 4 ( $p$ -value < 0.001). Overall, our data consistently indicate that chronic alcohol consumption can extensively perturb liquid biopsy-based early-stage cancer detection by forcing the serum protein concentration profiles of ever-drinker cancer patients to mimic those of healthy individuals without cancer (Figure 4D). To the best of our knowledge, this is the first direct evidence that patient lifestyle can have an extensive influence on the performance of cancer detection by liquid biopsy [39, 40] and necessitates further analyses of serum protein signatures in larger cohorts.

### 3. Discussion

Herein, we presented a comprehensive analysis of serum protein signatures in a prospective cohort that involved 601 participants with or without three of the most lethal cancers (lung, pancreas, and colorectum). We identified ten distinct serum protein signatures in cancer patients, which covered expansive aspects of cancer biology, including sex, age, obesity, tobacco smoking, alcohol consumption, family history, immune response, angiogenesis, and extracellular matrix remodeling (Figure 1B and Figure 2A, B).

The biological relevance of these serum protein signatures was studied by a comprehensive metadata analysis (Figure 2A, B). Sex and tobacco smoking had the most influence on serum protein signatures, while family history had the least. Sex- and smoking-dependent variations in these signatures were in consistent with previous reports [22, 25, 26, 27, 28]. Although our serum protein signatures only pertained to 48 serum proteins studied in this cohort, investigators can also assess the membership of their novel serum protein of interest in each signature using the MATLAB-based SeroHet software (<https://github.com/hanjanlee21/SeroHet>). As signatures 2–5 were highly enriched in cancer patients with clinical stage as early as stage I (Figure 3A–D), further expansion of these signatures will be valuable in achieving liquid biopsy-based cancer screening in clinics.

Our comprehensive analysis on cancer-associated serum protein signatures also provides investigators with resource to efficiently design high-performance liquid biopsy panels. Due to the scarcity of information on how serum proteins are regulated in cancer patients, rational design of serum protein panels has long been a formidable task for cancer biologists [9, 41]. Since we have identified multiple distinct signatures that are highly diagnostic of early-stage cancers, investigators may efficiently improve the performances of pre-existing blood tests by adding serum proteins that belong to signatures that are underrepresented in the previous liquid biopsy panel. In this sense, our study lays a practical foundation for signature-based rational design of high-performance liquid biopsies.

Importantly, based on an extensive metadata analysis, we discovered that the detection of cancer in patients with the history of chronic alcohol consumption is much harder compared to patients without the history of chronic alcohol consumption, as the serum protein concentration profiles of the former strongly phenocopy those of healthy individuals without cancer. The observed decrease in the expression of signatures 3 and 4 upon chronic alcohol intake (Figure 2A) was specific to cancer patients, as alcohol consumption exhibited less pronounced effects on the expression of these signatures in healthy individuals without cancer (Figures S6, S7). Downregulation of proinflammatory cytokines, such as IL-8 and TNF $\alpha$ , upon chronic alcohol consumption has been previously demonstrated [42]. By contrast, signature 6 showed an identical trend in both healthy individuals and cancer patients, with chronic alcohol consumption being positively correlated with the overexpression of signature 6 in both groups of participants (Figure 2A, Figure S6). This is in agreement with previous reports, which demonstrated upregulation of ferritin and ALDH1A1 upon chronic alcohol consumption [26, 27, 28]. Overall, our data indicate that there is both universal and cancer-specific regulation of serum protein signatures by chronic alcohol uptake, of which combined effects result in high resemblance of serum protein signatures of cancer patients with the history of alcohol consumption to those of healthy individuals without cancer. Examination of the history of chronic alcohol uptake enhanced the sensitivity of liquid biopsy in all stages, providing the direct evidence that patient lifestyle can affect the performance of blood tests. However, the inclusion of the new binary parameter on the history of alcohol consumption did not completely abolish the inequality in the performance of liquid biopsies, suggesting the need to develop better tools to identify cancers in ever-drinkers.

One potential approach is to implement abstinence from alcohol before each blood test. Acute alcohol intake is known to impair the release of proinflammatory cytokines, such as IL-6 and TNF $\alpha$ , in response to inflammatory stimuli [43], of which effect is known to persist days after administration. Since many liquid biopsy panels rely on the inflammatory response to the tumor, acute alcohol intake can have detrimental effects on the detection of cancer by these blood tests. However, it remains to be studied whether acute alcohol intake can indeed impair the performance of serum protein-based liquid biopsies, as our study instead focused on chronic alcohol consumption in cancer patients. Since pre-biopsy lifestyle intervention is a commonly utilized strategy to manage cancer in clinics [44], short-term alcohol intervention, such as a pre-analytical week-long cessation [43], is a viable option to enhance the performance of liquid biopsies. Another approach is to identify novel serum protein signatures that are differentially regulated in ever-drinker cancer patients compared to healthy individuals. Inclusion of the members of these signatures may significantly enhance the performance of serum protein-based liquid biopsies in detecting early-stage cancers. However, whether application of these approaches can indeed boost the performance of serum protein-based liquid biopsies await further clinical trials.

#### 3.1. Limitations of the study

One major downfall of the study is the relatively small size of the cohort that we have utilized to uncover serum protein signatures in

cancer patients. Rather, our study focused on increasing the number of serum proteins and types of metadata to ensure comprehensive analysis on serum protein signatures. Overall, our study involved 28,848 independent measurements on forty-eight different serum protein markers, which to date is one of the widest serum protein panels utilized in multicancer cohorts [11]. Another major downfall is the inherent bias in the selection of healthy individuals. Other implications other than cancer, such as viral infection, can modulate serum protein signatures and potentially affect the performance of liquid biopsies, but its validation requires further investigations. Finally, there were correlation of clinical parameters that were inherent to the population. For instance, female individuals in our cohort were an order of magnitude less likely to have a history of tobacco smoking compared to male individuals ( $p$ -value < 0.001, Fisher's exact test), which was a trend that is inherent to the population that we studied. Nevertheless, this study provides a comprehensive view on the regulation of serum proteins in cancer patients and provides the first evidence that patient lifestyle can strongly perturb the practice of cancer screening. We envision these observations to be invaluable in deciphering the role of serum proteins in cancer, a second leading cause of death for humankind.

## 4. Materials and methods

### 4.1. Study design

The current study included 601 East Asian individuals with or without three of the most lethal types of cancer worldwide: lung, pancreas, and colorectum. To ensure the representability of the cohort, every patient was re-evaluated of the primary origin of the tumor. 15 out of 480 cancer patients were later determined as having a non-primary tumor, of which protein concentration profiles we excluded for further investigations. After the participants provided written informed consent, peripheral blood samples as well as patient demographic and clinical data were prospectively collected. The participants did not consume anything but water 8–10 h before the fasting blood test. Peripheral blood samples were sent to the technicians as cryopreserved vials containing 500  $\mu$ L of serum. The biospecimen and data used in the study was provided by the Seoul National University Bundang Hospital Human Bioresource Center. Participants were diagnosed with hypertension if the systolic blood pressure was greater than or equal to 140 mmHg or the diastolic blood pressure was greater than or equal to 90 mmHg, following repeated examination ( $n \geq 2$ ). Participants were diagnosed with diabetes if either the fasting plasma glucose level was greater than or equal to 126 mg/dL, the casual glucose level was greater than or equal to 200 mg/dL, or the hemoglobin A1c level was greater than or equal to 6.5%. Obesity was diagnosed if the body mass index was greater than or equal to 25 kg/m<sup>3</sup>, per Asians of the Regional Office for the Western Pacific Region of the World Health Organization criteria [45]. Cancer staging was based on an 8th edition of the AJCC manual. Detailed information regarding the cohort is shown in Table S1. This study was approved by the Institutional Review Board of Seoul National University Bundang Hospital (X-2001-586-906) and compiled with the Health Insurance Portability and Accountability Act.

### 4.2. Quantification of serum proteins

The Luminex 200 system with xPONENT 4.3 (Luminex, Austin TX) was utilized to determine the concentration of each protein marker within the blood samples of individuals. Luminex bead-based multiplex immunoassays (Millipore, Bilerica NY) were split into a total of six panels: human angiogenesis/growth factor panel 1 (Millipore, Bilerica NY; catalog number, HANG2MAG-12K; angiopoietin-2, G-CSF, endoglin, FGF1, follistatin), human angiogenesis panel 2 (Millipore, Bilerica NY; catalog number, HAGP1MAG-12K; sAXL, sHER2, sE-selectin, TSP2, sEGFR, suPAR, sVEGFR1, sPECAM-1, OPN), human cancer/metastasis

biomarker panel 1 (Millipore, Bilerica NY; catalog number, HCMBMAG-22K; GDF15, DKK1, enolase 2, OPG), human circulating cancer biomarker panel 1 (Millipore, Bilerica NY; catalog number, HCCBP1MAG-58K; CA15-3, CA19-9, MIF, leptin, IL-6, CEA, IL-8, HGF, sFas, TNF $\alpha$ , PRL, SCF, Cyfra21-1, FGF2,  $\beta$ -HCG, HE4, TGF- $\alpha$ , VEGF), human circulating cancer biomarker panel 3 (Millipore, Bilerica NY; catalog number, HCCBP3MAG-58K; galectin-3, myeloperoxidase, SHBG, IGFBP3, ferritin), and human circulating cancer biomarker panel 4 (Millipore, Bilerica NY; catalog number, HCCB4MAG-58K; mesothelin, midkine, kallikrein 6, ALDH1A1, EpCAM, CD44). Less than 25  $\mu$ L of serum was used per panel for the bead-based multiplex immunoassay. Multiplex immunoassay procedures were performed according to manufacturer's instructions. Briefly, each well was prewashed by filling with 100  $\mu$ L of washing buffer for 10 min. After removal of washing buffer, 25  $\mu$ L of each standard, quality control, and assay buffer were added to appropriate wells, followed by the addition of 25  $\mu$ L of matrix solution. 25  $\mu$ L of mixed beads were added to each well and incubated on a plate shaker overnight at 4  $^{\circ}$ C. The plate was washed with 200  $\mu$ L of washing buffer 2 times and then 25  $\mu$ L of detection antibodies were added and incubated on plate shaker for 1 h at room temperature. 25  $\mu$ L of streptavidin-phycoerythrin were added and incubated on plate shaker for 30 min at room temperature and then washed with 200  $\mu$ L of washing buffer 2 times. 100  $\mu$ L sheath fluid were added to all wells, and the beads were resuspended on a plate shaker for 5 min at room temperature. The level of human serum amyloid A1 was determined by a solid phase sandwich enzyme-linked immunosorbent assay (ELISA) using a human serum amyloid A1 DuoSet ELISA kit (R&D Systems, Minneapolis MN; catalog number, DY3019-05). Concentration of each marker was determined after performing 5-parameter curve fits on logarithmized data using SoftMax Pro (version 5.4). All measurements were made within eight batches of immunoassay sessions, of which batch identifiers were documented in Table S3.

### 4.3. Hierarchical clustering

Hierarchical clustering was performed on OriginPro (version 2020b). Group average-clustering based on Euclidean distances was performed. Each parameter was mean-centered normalized before analysis, and the z-scores for each parameter were color-coded in the final heatmap. Individual metadata have been binarized and mean-centered normalized before hierarchical clustering.

### 4.4. Statistics

All statistical analyses were performed on OriginPro, if not explicitly stated otherwise. A two-tailed Welch's  $t$ -test was used when comparing the difference between two groups, as each group exhibited unequal variance. To assess the statistical significance of a given sensitivity at 90% specificity, a two-tailed one-sample Student's  $t$ -test was used against the baseline sensitivity of 10%. Spearman's correlation coefficient was used when assessing the correlation between two profiles of serum protein concentrations. ROC curves were drawn on OriginPro, with score and state values being generated from the SeroHet software, which performs 1,000 iterations of 10-fold cross-validation. Logistic regression was performed using the fitglm function of MATLAB assuming the logistic binomial model of probability. Binary parameter for chronic alcohol consumption was set as one if the individual has a positive history of chronic alcohol intake and as zero if the individual has a negative history of chronic alcohol intake. Bland-Altman test was used to derive  $p$ -values from 95% confidence intervals and was performed as previously described [46]. Signature-wise Bonferroni correction was utilized to assess the statistical significance of fold changes for each serum protein marker. \* $p$ -value < 0.05; \*\* $p$ -value < 0.01; and \*\*\* $p$ -value < 0.001 were considered statistically significant.

## Declarations

### Author contribution statement

Hanjun Lee: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Seo Yihl Kim; Dong wook Kim: Performed the experiments; Contributed reagents, materials, analysis tools or data.

Young Soo Park: Contributed reagents, materials, analysis tools or data.

Jin-Hyeok Hwang; Sukki Cho: Conceived and designed the experiments; Analyzed and interpreted the data.

Je-Yoel Cho: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

### Funding statement

Professor Je-Yoel Cho and Jin-Hyeok Hwang were supported by the Ministry of Science and ICT, South Korea [NRF-2021R1A5A1033157 & NRF-2017R1A2B2012191].

### Data availability statement

All original codes are available at <https://github.com/hanjunlee21/SeroHet> and archived at <https://doi.org/10.5281/zenodo.4898421>. All figures introduced in this study can be reproduced using the SeroHet software.

Any additional information regarding the study is available from the lead contact.

### Declaration of interest's statement

The authors declare no competing interests.

### Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2022.e12359>.

### Acknowledgements

We thank S. Yum for assistance in the management of clinical data prospectively collected from lung cancer patients. We are grateful to J.-M. An, J.Y. Kim, and E.-H. Yeon for their technical assistance. We also thank M.S. Lawrence for his critical review of the manuscript.

## References

- [1] A. Haber, Daniel, S. Gray, Nathanael, J. Baselga, The evolving war on cancer, *Cell* 145 (2011) 19–24.
- [2] B.-Z. Qian, J.W. Pollard, Macrophage diversity enhances tumor progression and metastasis, *Cell* 141 (2010) 39–51.
- [3] K.E. De Visser, A. Eichten, L.M. Coussens, Paradoxical roles of the immune system during cancer development, *Nat. Rev. Cancer* 6 (2006) 24–37.
- [4] D. Hanahan, A. Weinberg, Robert, Hallmarks of cancer: the next generation, *Cell* 144 (2011) 646–674.
- [5] S.R. Menakuru, N.J. Brown, C.A. Staton, M.W.R. Reed, Angiogenesis in pre-malignant conditions, *Br. J. Cancer* 99 (2008) 1961–1966.
- [6] D. Hanahan, J. Folkman, Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis, *Cell* 86 (1996) 353–364.
- [7] Z. Yurkovetsky, S. Skates, A. Lomakin, B. Nolen, T. Pulsipher, F. Modugno, J. Marks, A. Godwin, E. Gorelik, I. Jacobs, et al., Development of a multimarker assay for early detection of ovarian cancer, *J. Clin. Oncol.* 28 (2010) 2159–2166.
- [8] R.E. Brand, B.M. Nolen, H.J. Zeh, P.J. Allen, M.A. Eloubeidi, M. Goldberg, E. Elton, J.P. Arnoletti, J.D. Christein, S.M. Vickers, et al., Serum biomarker panels for the detection of pancreatic cancer, *Clin. Cancer Res.* 17 (2011) 805–816.
- [9] M. Kälin, I. Cima, R. Schiess, N. Fankhauser, T. Powles, P. Wild, A. Templeton, T. Cerny, R. Aebersold, W. Krek, Novel prognostic markers in the serum of patients with castration-resistant prostate cancer derived from quantitative analysis of the pten conditional knockout mouse proteome, *Eur. Urol.* 60 (2011) 1235–1243.
- [10] C. Wingren, A. Sandstrom, R. Segersvard, A. Carlsson, R. Andersson, M. Lohr, C.A.K. Borrebaeck, Identification of serum biomarker signatures associated with pancreatic cancer, *Cancer Res.* 72 (2012) 2481–2490.
- [11] J.D. Cohen, L. Li, Y.X. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A.A. Javed, F. Wong, A. Mattox, et al., Detection and localization of surgically resectable cancers with a multi-analyte blood test, *Science* 359 (2018) 926.
- [12] J.D. Cohen, A.A. Javed, C. Thoburn, F. Wong, J. Tie, P. Gibbs, C.M. Schmidt, M.T. Yip-Schneider, P.J. Allen, M. Schattner, et al., Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers, *Proc. Natl. Acad. Sci. U.S.A.* 114 (2017) 10202–10207.
- [13] T. Jiang, S.X. Ren, C.C. Zhou, Multi-cancer blood testing combined with PET-CT: road for hope to screen for cancer and guide intervention, *Signal Transduct. Targeted Ther.* 5 (2020) 2.
- [14] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2020, *CA Cancer J. Clin.* 70 (2020) 7–30.
- [15] H. Graham, D.J. Chandler, S.A. Dunbar, The genesis and evolution of bead-based multiplexing, *Methods* 158 (2019) 2–11.
- [16] Y.F. Ma, Y. Ren, Z.J. Dai, C.J. Wu, Y.H. Ji, J.R. Xu, IL-6, IL-8 and TNF-alpha levels correlate with disease stage in breast cancer patients, *Adv. Clin. Exp. Med.* 26 (2017) 421–426.
- [17] S.A. Yoo, L. Leng, B.J. Kim, X. Du, P.V. Tilstam, K.H. Kim, J.S. Kong, H.J. Yoon, A.H. Liu, T. Wang, et al., MIF allele-dependent regulation of the MIF coreceptor CD44 and role in rheumatoid arthritis, *Proc. Natl. Acad. Sci. U.S.A.* 113 (2016) E7917–E7926.
- [18] M.J. Cross, L. Claesson-Welsh, FGF and VEGF function in angiogenesis: signalling pathways, biological responses and therapeutic inhibition, *Trends Pharmacol. Sci.* 22 (2001) 201–207.
- [19] P. Bornstein, L.C. Armstrong, K.D. Hankenson, T.R. Kyriakides, Z.T. Yang, Thrombospondin 2, a matricellular protein with diverse functions, *Matrix Biol.* 19 (2000) 557–568.
- [20] B. Hu, P. Guo, Q. Fang, H.Q. Tao, D.G. Wang, M. Nagane, H.J.S. Huang, Y. Gunji, R. Nishikawa, K. Alitalo, et al., Angiopoietin-2 induces human glioma invasion through the activation of matrix metalloprotease-2, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 8904–8909.
- [21] B. He, M. Mirza, G.F. Weber, An osteopontin splice variant induces anchorage independence in human breast cancer cells, *Oncogene* 25 (2006) 2192–2202.
- [22] H.J. Guйда, H.G. Friesen, Serum prolactin levels in humans from birth to adult life, *Pediatr. Res.* 7 (1973) 534–540.
- [23] J. Terkel, C.A. Blake, V. Hoover, C.H. Sawyer, Pup survival and prolactin levels in nicotine-treated lactating rats, *Proc. Soc. Exp. Biol. Med.* 143 (1973) 1131–1135.
- [24] K. Andersson, P. Eneroth, K. Fuxe, F. Mascagni, L.F. Agnati, Effects of chronic exposure to cigarette-smoke on amine levels and turnover in various hypothalamic catecholamine nerve-terminal systems and on the secretion of pituitary-hormones in the male-rat, *Neuroendocrinology* 41 (1985) 462–466.
- [25] K. Hakkinen, A. Pakarinen, Muscle strength and serum testosterone, cortisol and SHBG concentrations in middle-aged elderly men and women, *Acta Physiol. Scand.* 148 (1993) 199–207.
- [26] M. Patel, L. Lu, D.S. Zander, L. Sreerama, D. Coco, J.S. Moreb, ALDH1A1 and ALDH3A1 expression in lung cancers: correlation with histologic type and potential precursors, *Lung Cancer* 59 (2008) 340–349.
- [27] S.H. Lee, E.Y. Kim, S.K. Lee, J.Y. Hong, Y.S. Kim, The relationship between serum ferritin levels, smoking, and lung function in Korean: analysis of the Fourth Korea National Health and Nutrition Examination Survey (KNHANES IV), *Chest* 140 (2011) 568A.
- [28] A. Rossi, M. Voigtlaender, H. Klose, H. Schluter, G. Schon, S. Loges, M. Paolini, C. Bokemeyer, M. Reck, G. Tarro, M. Binder, High aldehyde dehydrogenase levels are detectable in the serum of patients with lung cancer and may be exploited as screening biomarkers, *J. Oncol.* 11 (2019), 8970645.
- [29] H. Kristenson, G. Fex, E. Trelle, Serum ferritin, gammaglutamyl-transferase and alcohol-consumption in healthy middle-aged men, *Drug Alcohol Depend.* 8 (1981) 43–50.
- [30] N. Milman, M. Kirchoff, Relationship between serum ferritin, alcohol intake, and social status in 2235 Danish men and women, *Ann. Hematol.* 72 (1996) 145–151.
- [31] A.W.K. Chan, Racial-differences in alcohol sensitivity, *Alcohol Alcohol.* 21 (1986) 93–104.
- [32] A. Yoshida, V. Dave, R.J. Ward, T.J. Peters, Cytosolic aldehyde dehydrogenase (ALDH1) variants found in alcohol flushers, *Ann. Hum. Genet.* 53 (1989) 1–7.
- [33] O. Tredan, M. Manuel, G. Clapissou, T. Bachelot, S. Chabaud, C. Bardin-dit-Courageot, C. Rigal, C. Biota, A. Bajard, N. Pasqual, et al., Patients with metastatic breast cancer leading to CD4(+) T cell lymphopenia have poor outcome, *Eur. J. Cancer* 49 (2013) 1673–1682.
- [34] J.N. Hofmann, O. Landgren, R. Landy, T.J. Kemp, L. Santo, C.M. McShane, J.J. Shearer, Q. Lan, N. Rothman, L.A. Pinto, et al., A prospective study of circulating chemokines and angiogenesis markers and risk of multiple myeloma and its precursor, *JNCI Cancer Spectr.* 4 (2020) 9, pkz104.
- [35] W.H. Chow, G. Gridley, J.F. Fraumeni, B. Jarvholm, Obesity, hypertension, and the risk of kidney cancer in men, *N. Engl. J. Med.* 343 (2000) 1305–1311.
- [36] S.Y. Lee, M.T. Kim, S.H. Jee, J.S. Im, Does hypertension increase mortality risk from lung cancer? A prospective cohort study on smoking, hypertension and lung cancer risk among Korean men, *J. Hypertens.* 20 (2002) 617–622.
- [37] A. Lindgren, E. Pukkala, A. Nissinen, J. Tuomilehto, Blood pressure, smoking, and the incidence of lung cancer in hypertensive men in North Karelia, Finland, *Am. J. Epidemiol.* 158 (2003) 442–447.
- [38] E. Giovannucci, D.M. Harlan, M.C. Archer, R.M. Bergenstal, S.M. Gapstur, L.A. Habel, M. Pollak, J.G. Regensteiner, D. Yee, Diabetes and Cancer A consensus report, *Diabetes Care* 33 (2010) 1674–1685.

- [39] D. Belpomme, P. Irigaray, A.J. Sasco, J.A. Newby, V. Howard, R. Clapp, L. Hardell, The growing incidence of cancer: role of lifestyle and screening detection (Review), *Int. J. Oncol.* 30 (2007) 1037–1049.
- [40] Y. Balavarca, K. Weigl, H. Thomsen, H. Brenner, Performance of individual and joint risk stratification by an environmental risk score and a genetic risk score in a colorectal cancer screening setting, *Int. J. Cancer* 146 (2020) 627–634.
- [41] I. Cima, R. Schiess, P. Wild, M. Kaelin, P. Schuffler, V. Lange, P. Picotti, R. Ossola, A. Templeton, O. Schubert, et al., Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 3342–3347.
- [42] S. Arbabi, I. Garcia, G.J. Bauer, R.V. Maier, Alcohol (ethanol) inhibits IL-8 and TNF: role of the p38 pathway, *J. Immunol.* 162 (1999) 7441–7445.
- [43] N.B.D. El-Guindy, W.J. de Villiers, D.E. Doherty, Acute alcohol intake impairs lung inflammation by changing pro- and anti-inflammatory mediator balance, *Alcohol* 41 (2007) 335–345.
- [44] J.S. Ostroff, J.E. Burkhalter, P.M. Cinciripini, Y.L. Li, M.P. Shiyko, C.Y. Lam, J.L. Hay, L.K. Dhingra, J. Lord-Bessen, S.M. Holland, R. Manna, Randomized trial of a presurgical scheduled reduced smoking intervention for patients newly diagnosed with cancer, *Health Psychol.* 33 (2014) 737–747.
- [45] W.H. Organization, *The Asia-Pacific Perspective: Redefining Obesity and its Treatment*, 2000.
- [46] D.G. Altman, J.M. Bland, How to obtain the P value from a confidence interval, *BMJ* 343 (2011) d2304–d2304.