Pan-cancer DNA methylation signature quantification of lifestyle exposures and cancer prognosis

Kangpei Tao
Imperial College London
Jaim Sutton
Imperial College London
James M. Flanagan (j.flanagan@imperial.ac.uk)
Imperial College London

Research Article

Keywords: TCGA, epigenome-wide association study, epigenetics, DNA methylation, signatures, BMI, alcohol, smoking, cancer, survival, cancer risk

Posted Date: February 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-199381/v1

Abstract

Background: Alcohol consumption, body mass index (BMI) and cigarette smoking are among the most well-studied lifestyle cancer risk exposures which can also change the host's epigenetic methylation patterns. Some of the changes associated with lifestyle exposure are specific and stable over time, thus, can be used to predict and quantify the exposure. Although the link between these lifestyle exposures and increased odds ratio (OR) of different cancer types is well known, their role in predicting cancer survival remains less clear. We hypothesized that by using predicted lifestyle exposures based on the methylation profiles in tumour DNA we could predict the overall survival probability in cancer patients associated with these exposures.

Results: The Cancer Genome Atlas (TCGA) Pan-Cancer dataset was used to test the prognostic value of the predicted DNAmethylation (DNAm) alcohol, BMI and smoking exposures in 24 cancer types (n= 8,238 subjects). Multivariable Cox proportional hazards models with adjustment for age, cancer stage and other exposures were used to calculate the hazards ratio (HR) for overall survival associated with these predicted DNAm exposures. We observed specific cancer types with strong associations between poorer survival and higher alcohol consumption (bladder, brain, esophageal, and head and neck cancers), higher BMI (bladder, pancreatic and post-menopausal breast cancers), and smoking (B-cell lymphoma, stomach, bladder and lung cancers). Interestingly, we also observed associations between better survival from kidney cancer with higher alcohol consumption and smoking exposures. For alcohol consumption we found a positive association between HR and OR across all cancers, indicating that for cancers where alcohol is a significant risk factor, it is also associated with poorer survival ($p = 0.022$). This was not the case for the BMI ($p = 0.548$) or smoking exposures ($p = 0.193$).

Conclusions: In conclusion, these DNAm exposure signatures may provide novel information on the relationship between these lifestyle factors and cancer outcomes.

Introduction

Obesity, alcohol and smoking consumption are among the most studied lifestyle exposures known to be associated with increased cancer risk for many cancer types (1–5). However, less is known about the value of these lifestyle factors in predicting cancer patient's survival. Epidemiological studies have reported that higher alcohol consumption is associated with poorer esophageal, head and neck, pancreatic and colorectal cancer survival (6–9); higher BMI is associated with poorer breast, ovarian, pancreatic, bladder and colorectal cancer survival (10–15); and smoking is associated with poorer lung, B-cell lymphoma, stomach and bladder cancer outcomes (16–19). These studies used reported clinical or questionnaire exposure data to analyse their association with prognosis. However, this reported exposure data is typically captured by questionnaires, which can be subjected to measurement error, recall bias and patient underestimation. Furthermore, these studies frequently only analysed one cancer type at a time which makes inter-cancer type comparisons difficult.

DNA methylation is a commonly studied epigenetic modification characterised by the addition of a methyl group to DNA, typically at a cytosine-phosphate-guanine (CpG) nucleotide base pairing. These modifications have been shown to be dynamic and stable, tissue and cell-specific, involved in transcription and gene regulation, and can be influenced by genetic, demographic and lifestyle exposures (20,21). Many epigenome-wide association studies (EWAS) and meta-analyses have identified DNAm signatures that are associated with lifestyle exposures that encompass genome-wide CpG methylation differences in the extreme levels of each exposure, compared to those without the exposure (22–25). Additionally, age acceleration, BMI, alcohol, smoking and estrogen DNAm exposure signatures have been found to be associated with breast and lung cancer risk (26–31). These studies usually only measure DNAm lifestyle exposures in patients' blood and have not investigated if DNAm exposures are associated with prognosis. However DNAm signatures in blood can be observed in other tissues and cells (26,32,33), and the TCGA Pan-Cancer dataset has patient's DNA methylation and clinical information on over 30 cancer types that can be used for DNAm exposure associated survival comparisons in multiple cancers.

In this study, we therefore hypothesised that published alcohol, BMI and smoking DNAm exposure signatures can be used to measure lifestyle exposures in tumour DNA and then used to predict patients overall survival probability in different types of cancer and are further related to cancer risk. Firstly for 24 TCGA cancer types we were able to extract the ORs for the association of these reported lifestyle exposures with cancer risk, from published meta-analyses (1–5,34–36). Next, we measured these lifestyle exposures using their respective published DNAm exposures signatures in DNA methylation data for these cancer types and available matched adjacent normal tissues, from the TCGA Pan-Cancer database. For each cancer type we then calculated the DNAm exposure associated HRs and then compared this to their literature reported exposure associated ORs, to further measure how well each exposure cancer survival correlated with their respective exposure cancer risk.

Methods

Study population and data

The TCGA collection, contains approximately 11,000 patient tissue samples, covering 33 cancer types to date and includes patient molecular assay datasets and clinical data. This collection has also been standardized for inter-cancer type comparisons with four major clinical endpoints and optimized for various omics studies into the TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR) (37). In this study, we first identified and reviewed published meta-analyses that contained information on reported BMI, alcohol and smoking consumption exposure associations with cancer risk. The reported exposure ORs (relative risks (RR) were also called ORs in this study) and confidence intervals (CIs) were then extracted from these studies for each cancer type that DNA methylation, clinical and survival data was also available for in the TCGA Pan-Cancer collection (1–5,34–36). In total this was for 24 cancer types, for which, patients age, tumour-node-metastasis (TNM) stage ('stages I and II' and 'stages III and V' were combined), overall survival times and vital status clinical data and Illumina methylation BeadChip beta-value data for their tumour and available adjacent normal tissue was obtained from the TCGA Pan-

Cancer datasets stored on the TCGA University of California Santa Cruz (UCSC) Xena browser (https://xenabrowser.net). For these patients, 8,238 had DNA methylation data available for their primary tumour tissue, 722 for their adjacent normal tissue, and 696 of these were from both tissues. The TCGA 450K methylation beta-value data was previously pre-processed through standard quality control steps such as probe filtering, normalisation, and correction for batch effects using the minfi package (38).

DNAm exposure signature measurements

The DNAm exposure signatures used in this study, were obtained from previously published EWASs measuring overweight/obese BMI, moderate to heavy alcohol consumption and current smoking behaviour exposures (25). These DNAm exposure signatures where then used to predict these lifestyle exposures in patients primary tumour and available adjacent normal tissue from the pre-processed TCGA 450K methylation data. For each of these DNAm exposure signatures the number of CpG sites and methylation beta-values that were available (due to missing CpGs in the pre-processed TCGA methylation data) were as follows: 612/1109, 262/450 and 132/233 for BMI, alcohol and smoking respectively. For each exposure and patient, a DNAm score was calculated from the total sum of each of their exposure CpG site beta-values multiplied by their corresponding DNAm exposure signature beta coefficients. Patients DNAm exposure scores were then standardized to z-scores within the study cohort, for subsequent inter-cancer comparison. A summary of the study patients clinical data and predicted DNAm alcohol, BMI and smoking exposures for each cancer type is available in Supplementary Table 1.

DNAm exposures in tumour and adjacent normal tissue comparison

The predicted DNAm exposures derived from patients adjacent normal tissue was used to assess the performance of the DNAm exposure signatures performance in the primary tumour tissues, to investigate whether tumour DNA represented an accurate representation of the exposure as determined in the normal tissue. Firstly, to compare the consistency of the DNAm exposures measurements between tumour and adjacent normal tissue, Spearman's rank correlation coefficients were calculated for each of the DNAm exposure signatures CpG sites beta-values between the patients tumour and matched adjacent normal tissue (n=696 patients). These correlation coefficients were then normalized into proportional frequencies by dividing by the number of total CpG sites associated with each exposure. The distribution of these CpG site tumour versus normal correlation coefficients were then plotted for each exposure, against their proportional frequencies. Next, for each exposure the patients tumour versus normal correlation coefficients were plotted against their DNAm exposure signature CpG site beta coefficients (impact of each CpG site in the DNAm exposure signatures), to examine each of the CpG sites relationship between these two tissues according to their weights of contribution. For each exposure, hierarchical cluster analysis using the Manhattan distance was then carried out for the patients tumour and adjacent normal tissue DNA exposure CpG beta-values and visualised in dendrograms, to see how similar the DNAm exposure measurements were in the two tissue types. Lastly, Pearson's correlation coefficients were calculated between the predicted DNAm exposure z-scores in patients tumour and matched adjacent normal tissue.

Exposures, cancer prognosis and risk

DNAm exposure survival analysis was then carried out by univariable and multivariable Cox proportional hazards analysis, by calculating the HR associated with each DNAm exposure z-score for each cancer type. The DNAm exposure multivariable HR models were adjusted for age at diagnosis, TNM stage (where available) and relevant DNAm exposure scores, for each cancer type. Where, the DNAm alcohol and BMI exposure associated HR models were adjusted for DNAm smoking, and the DNAm smoking associated HR models for DNAm BMI, as these DNAm exposures were found to significantly confound the analyses. Additionally the breast cancer data was also stratified into pre and post-menopausal breast cancers due to the well-known association between BMI and menopausal status at time of diagnosis (1). These breast cancer subgroups were then also analysed for DNAm BMI associated survival, with adjustment for age at diagnosis, TNM stage (where available) and DNAm smoking scores. Next, to test the association between the DNAm exposure cancer survival and reported exposure cancer risk, for each exposure and cancer type the log-transformed multivariable DNAm exposure HRs were regressed on their respective log-transformed reported exposure ORs for each cancer type by linear regression, after further adjustment for the group size of each cancer type. The group size was normalized by division of the largest group size in the studied cancer types. All statistical analyses in this study were carried out using R version 3.6.0.

Results

DNAm exposures in tumour and normal tissue

For all the DNAm exposures, the majority of the associated CpG sites beta-values were moderately and significantly correlated between the patients tumour and adjacent normal tissue, indicating that the exposures were similarly affecting tumour and normal tissue within each individual. The normalised distribution of the patients tumour versus adjacent normal correlation coefficients for each DNAm exposure are shown in Fig. 1A. For the DNAm alcohol, BMI and smoking exposure CpG sites, the median of the Spearman's correlation coefficients was 0.35 (interquartile range (IQR): 0.27–0.44), 0.36 (IQR: 0.27–0.45) and 0.35 (IQR): 0.27–0.44) respectively. Furthermore, there was no relationship between the DNAm exposure signature CpG site beta coefficients and their correlation coefficients between the patients tumour and normal tissues (Fig. 1B). After the hierarchical clustering analysis, the DNAm alcohol and BMI exposure signature CpG beta-values did not have a tendency to separate into normal and tumour tissue samples, indicating that these groups did not cluster separately (Figs. 2A, 2B). While, the DNAm smoking exposure CpG beta values did have more of a tendency to separate into normal and tumour tissue samples, indicating the existence of a systematic difference between the smoking exposure signature beta-values in both tissue types (Fig. 2C). DNAm alcohol and BMI exposure z-scores in patients tumour and adjacent normal tissue were moderately correlated, with Pearson's correlation coefficients of 0.55 and 0.39 respectively (Figs. 3A and 3B). The DNAm smoking exposure z-scores in patients tumour and adjacent normal tissue was weakly correlated, with a correlation coefficient of 0.2 (Fig. 3C). This is

consistent with the hierarchical clustering analysis findings, indicating that tumour and normal tissue methylation levels were different for the DNAm smoking exposure CpG sites.

DNAm exposures and cancer prognosis

The DNAm exposure associated HR analyses and reported exposures associated ORs for each of the 24 cancer types for cancer survival and risk respectively, are shown in Tables 1–3 with model fit statistics presented in Supplementary Table 2. These tables show that unadjusted high DNAm alcohol exposures were significantly associated with poorer survival in patients with bladder (BLCA) and brain (LGG) cancers and with better survival in thyroid (THCA) and kidney (KIRC, KIRP) cancers. However, after adjustment for potential confounding factors, high DNAm alcohol exposures remained significantly associated with survival in BLCA (HR = 1.3, 95% CI (1.0–1.6), p = 0.020), KIRP (HR = 0.47, 95% CI (0.3–0.8), p < 0.01), LGG (HR = 1.6, 95% CI (1.3–2.0), p < 0.0001), and became significantly associated with poorer survival in esophageal (ESCA) (HR = 1.5, 95% CI (1.0–2.1), p = 0.030), and head and neck (HNSC) (HR = 1.3, 95% CI (1.0–1.6), p = 0.042) cancers. The unadjusted high DNAm BMI exposures were significantly associated with poorer survival in patients with bladder (BLCA), postmenopausal breast (BRCA), brain (LGG), pancreatic (PAAD) and rectal (READ) cancers and with better survival in kidney (KIRC) cancer. However, after adjustment for potential confounding factors, high DNAm BMI exposures remained significantly associated with survival in BLCA (HR = 1.2, 95% CI (1.0–1.4), p = 0.015), postmenopausal BRCA (HR = 1.44, 95% CI (1.1–2.0), p = 0.018) and PAAD (HR = 2.1, 95% CI (1.5–3.0), p < 0.0001) cancers. The unadjusted DNAm smoking exposures were significantly associated with poorer survival in patients with B-cell lymphoma (DLBC), lung (LUSC) and stomach (STAD) cancers and better survival in kidney (KIRC, KIRP) cancers. However, after adjustment for potential confounding factors, the DNAm smoking exposures remained significantly associated with survival in DLBC (HR = 2.7, 95% CI (1.1–6.5), p = 0.028), KIRC (HR = 0.7, 95% CI (0.54–0.9), p < 0.01), LUSC (HR = 1.2, 95% CI (1.0–1.5), p = 0.049), and STAD (HR = 1.3, 95% CI (1.1–1.7), p = 0.016) cancers and became significantly associated with poorer survival in bladder cancer (BLCA) (HR = 1.2, 95% CI (1.0–1.5), p = 0.034). For 14 out of 24 tumour types we were also able to adjust for response to first line treatment (complete response versus stable disease/progression) as a potential confounder and found the majority of results remained the similar (Supplementary Table 3).

Table 1

Summary of DNAm alcohol exposure cancer survival and reported exposure cancer risks. For 24 cancer types, the DNAm alcohol exposure associated HRs for cancer survival were calculated from univariable and multivariable Cox proportional hazard model analysis and the reported alcohol exposure associated ORs for cancer risk were gathered from meta-analyses in the literature. DNAm alcohol z-scores were used as the dependent variable in the univariable and multivariable Cox proportional hazards models, with the multivariable HR analysis also adjusted for age at diagnosis, TNM stage (where applicable) and DNAm smoking exposure scores.

Alcohol							
Cancer	Cancer Risk	Univariable - survival			Multivariable - survival		
	OR (95% CI)	HR (95% CI)	P-value	N	HR (95% CI)	P-value	N
BLCA	0.95 (0.75–1.20)	**1.40 (1.10 - 1.70)**	**	409	**1.30 (1.00–1.60)**	**0.020**	407
BRCA	1.61 (1.33–1.94)	1.10 (0.82–1.40)	0.680	774	1.30 (0.97–1.70)	0.087	763
CESC	0.90 (0.73–1.11)	1.20 (0.91–1.50)	0.230	299	1.10 (0.86–1.40)	0.400	292
CHOL	2.64 (1.62–4.30)	1.70 (0.81–3.70)	0.160	36	1.70 (0.70–4.10)	0.240	36
COAD	1.44 (1.25–1.65)	1.10 (0.76–1.70)	0.550	290	1.00 (0.67–1.60)	0.920	280
DLBC	0.75 (0.64–0.88)	2.40 (0.83–7.10)	0.100	47	1.20 (0.25–5.60)	0.840	41
ESCA	4.95 (3.86–6.34)	1.40 (1.00–1.90)	0.050	174	**1.50 (1.00–2.10)**	**0.030**	168
GBM	1.45 (0.69–3.08)	1.30 (0.92–1.70)	0.160	124	1.20 (0.88–1.70)	0.230	124
HNSC	5.13 (4.31–6.10)	1.20 (0.97–1.50)	0.086	523	**1.30 (1.00–1.60)**	**0.042**	523
KICH	0.79 (0.72–0.86)	0.57 (0.29–1.10)	0.100	66	0.79 (0.41–1.50)	0.500	66
KIRC	0.79 (0.72–0.86)	**0.54 (0.40–0.75)**	**	316	0.83 (0.60–1.20)	0.270	314
KIRP	0.79 (0.72–0.86)	**0.56 (0.36–0.86)**	*	274	**0.47 (0.29–0.76)**	*	256

Abbreviations: confidence interval (CI), deoxyribonucleic acid methylation (DNAm), number of patients (N), hazards ratio (HR), odds ratio (OR) and tumour-node-metastasis (TNM)

Significant HR associations are shown in bold

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

Alcohol							
LGG	1.45 (0.69–3.08)	**1.40 (1.10–1.80)**	*	504	**1.60 (1.30–1.90)**	***	504
LIHC	2.07 (1.66–2.58)	1.10 (0.89–1.30)	0.440	375	1.10 (0.92–1.40)	0.240	351
LUAD	1.15 (1.02–1.30)	0.84 (0.66–1.10)	0.150	455	0.79 (0.61–1.00)	0.064	451
LUSC	1.15 (1.02–1.30)	1.10 (0.89–1.40)	0.340	365	1.10 (0.87–1.40)	0.460	362
OV	1.03 (0.95–1.12)	1.50 (0.61–3.70)	0.370	10	7.10 (0.98–52.00)	0.052	10
PAAD	1.19 (1.11–1.28)	1.20 (0.83–1.70)	0.360	184	1.20 (0.81–1.60)	0.430	181
PRAD	1.09 (0.98–1.21)	1.50 (0.49–4.80)	0.460	484	1.90 (0.59–6.50)	0.280	484
READ	1.44 (1.25–1.65)	0.64 (0.26–1.60)	0.330	94	0.75 (0.19–2.90)	0.680	85
STAD	1.44 (1.25–1.65)	1.20 (0.96–1.60)	0.110	393	1.30 (0.96–1.60)	0.100	382
THCA	0.81 (0.71–0.94)	**0.41 (0.21–0.81)**	**0.011**	502	0.59 (0.23–1.50)	0.270	500
UCEC	0.99 (0.84–1.16)	1.00 (0.79–1.30)	0.830	425	1.10 (0.83–1.40)	0.560	425
UCS	1.33 (1.01–1.76)	1.00 (0.66–1.50)	0.980	57	1.10 (0.67–1.70)	0.790	57

Abbreviations: confidence interval (CI), deoxyribonucleic acid methylation (DNAm), number of patients (N), hazards ratio (HR), odds ratio (OR) and tumour-node-metastasis (TNM)

Significant HR associations are shown in bold

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

Table 2
Summary of DNAm BMI exposure cancer survival and reported exposure cancer risks.

BMI							
Cancer	Cancer Risk	Univariable - survival			Multivariable - survival		
	OR (95% CI)	HR (95% CI)	P-value	N	HR (95% CI)	P-value	N
BLCA	1.05 (0.99–1.12)	**1.30 (1.10–1.50)**	*	409	**1.20 (1.00–1.40)**	**0.015**	407
BRCA[1]	0.89 (0.85–0.94)	1.68 (0.87–3.24)	0.120	167	1.84 (0.95–3.55)	0.070	165
BRCA[2]	1.05 (1.03–1.08)	**1.56 (1.15–2.12)**	**0.004**	495	**1.44 (1.06–1.96)**	**0.018**	492
CESC	1.14 (1.03–1.26)	1.00 (0.78–1.30)	0.900	299	0.96 (0.73–1.30)	0.780	292
CHOL	1.50 (1.21–1.85)	1.00 (0.72–1.50)	0.840	36	1.10 (0.72–1.80)	0.570	36
COAD	1.11 (1.07–1.15)	0.98 (0.65–1.50)	0.910	290	0.86 (0.57–1.30)	0.470	280
DLBC	1.00 (0.95–1.05)	1.10 (0.40–2.80)	0.890	47	0.95 (0.38–2.40)	0.920	41
ESCA	1.16 (1.09–1.24)	1.10 (0.78–1.40)	0.730	174	1.10 (0.75–1.50)	0.770	168
GBM	1.02 (0.94–1.10)	1.30 (0.77–2.10)	0.340	124	0.82 (0.49–1.40)	0.450	124
HNSC	1.07 (0.91–1.26)	1.00 (0.86–1.20)	0.910	523	1.00 (0.89–1.20)	0.570	523
KICH	1.25 (1.13–1.38)	1.90 (0.73–5.10)	0.180	66	2.00 (0.67–5.80)	0.220	66
KIRC	1.25 (1.13–1.38)	**0.70 (0.53–0.93)**	**0.015**	316	0.84 (0.63–1.10)	0.260	314

Abbreviations: body mass index (body mass index), confidence interval (CI), deoxyribonucleic acid methylation (DNAm), number of patients (N), hazards ratio (HR), odds ratio (OR) and tumour-node-metastasis (TNM)

[1] premenopausal breast cancer

[2] postmenopausal breast cancer

Significant HR associations are shown in bold

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

BMI							
KIRP	1.25 (1.13–1.38)	0.97 (0.66–1.40)	0.900	274	0.91 (0.61–1.40)	0.650	256
LGG	1.02 (0.94–1.10)	**1.60 (1.10–2.30)**	**0.021**	504	1.10 (0.72–1.60)	0.740	504
LIHC	1.26 (1.14–1.40)	0.95 (0.78–1.20)	0.620	375	0.96 (0.77–1.20)	0.680	351
LUAD	0.99 (0.93–1.05)	1.00 (0.75–1.30)	1.000	455	0.96 (0.72–1.30)	0.800	451
LUSC	0.99 (0.93–1.05)	0.91 (0.72–1.10)	0.410	365	0.92 (0.72–1.20)	0.470	362
OV	1.08 (1.02–1.15)	3.70 (0.57–23.00)	0.170	10	12.00 (0.93–150)	0.057	10
PAAD	1.11 (1.03–1.19)	**2.00 (1.40–2.70)**	***	184	**2.10 (1.50–3.00)**	***	181
PRAD	0.96 (0.93–0.99)	0.81 (0.34–2.00)	0.650	484	0.76 (0.32–1.80)	0.540	484
READ	1.05 (0.99–1.12)	**5.40 (1.60–18.0)**	*	94	3.00 (0.78–11.00)	0.110	85
STAD	1.08 (1.00–1.18)	1.20 (0.91–1.50)	0.230	393	1.20 (0.91–1.50)	0.250	382
THCA	1.11 (0.99–1.25)	1.20 (0.67–2.10)	0.530	502	1.40 (0.74–2.60)	0.310	500
UCEC	2.98 (2.63–3.39)	0.89 (0.65–1.20)	0.480	425	1.00 (0.76–1.40)	0.860	425
UCS	1.63 (1.55–1.71)	1.00 (0.69–1.50)	0.880	57	1.20 (0.80–1.90)	0.340	57

Abbreviations: body mass index (body mass index), confidence interval (CI), deoxyribonucleic acid methylation (DNAm), number of patients (N), hazards ratio (HR), odds ratio (OR) and tumour-node-metastasis (TNM)

[1] premenopausal breast cancer

[2] postmenopausal breast cancer

Significant HR associations are shown in bold

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

Table 3

Summary of DNAm smoking exposure cancer survival and reported exposure cancer risks. For 24 cancer types, the DNAm smoking exposure associated HRs for cancer survival were calculated from univariable and multivariable Cox proportional hazard model analysis and the reported smoking exposure associated ORs for cancer risk were gathered from meta-analyses in the literature. DNAm smoking z-scores were used as the dependent variable in the univariable and multivariable Cox proportional hazards models, with the multivariable HR analysis also adjusted for age at diagnosis, TNM stage (where applicable) and DNAm BMI exposure scores.

Smoking							
Cancer	Cancer risk	Univariable - survival			Multivariable - survival		
	OR (95% CI)	HR (95% CI)	P-value	N	HR (95% CI)	P-value	N
BLCA	3.29 (2.61–4.15)	1.10 (0.96–1.30)	0.140	409	**1.20 (1.00–1.50)**	**0.034**	407
BRCA	1.13 (1.04–1.22)	0.95 (0.75–1.20)	0.640	774	0.96 (0.76–1.20)	0.740	763
CESC	2.03 (1.49–2.57)	1.10 (0.84–1.40)	0.490	299	1.10 (0.84–1.50)	0.460	292
CHOL	1.45 (1.11–1.88)	1.30 (0.64–2.80)	0.440	36	1.50 (0.65–3.20)	0.360	36
COAD	1.25 (1.14–1.37)	0.92 (0.69–1.20)	0.530	290	0.86 (0.63–1.20)	0.320	280
DLBC	1.16 (0.98–1.37)	**2.30 (1.10–4.70)**	**0.021**	47	**2.70 (1.10–6.50)**	**0.028**	41
ESCA	3.10 (2.68–3.58)	0.92 (0.73–1.20)	0.520	174	0.97 (0.74–1.30)	0.820	168
GBM	1.08 (0.94–1.25)	0.80 (0.56–1.20)	0.240	124	0.95 (0.65–1.40)	0.770	124
HNSC	4.83 (3.72–6.29)	0.97 (0.82–1.10)	0.730	523	1.00 (0.87–1.20)	0.680	523
KICH	2.10 (1.77–2.50)	0.82 (0.48–1.40)	0.480	66	0.99 (0.55–1.80)	0.980	66
KIRC	2.10 (1.77–2.50)	**0.58 (0.46–0.74)**	***	316	**0.70 (0.54–0.90)**	*	314

Abbreviations: body mass index (body mass index), confidence interval (CI), deoxyribonucleic acid methylation (DNAm), number of patients (N), hazards ratio (HR), odds ratio (OR) and tumour-node-metastasis (TNM)

Significant HR associations are shown in bold

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

Smoking							
KIRP	2.10 (1.77–2.50)	**0.68 (0.49–0.94)**	**0.021**	274	0.90 (0.62–1.30)	0.590	256
LGG	1.08 (0.94–1.25)	1.20 (0.89–1.70)	0.200	504	1.10 (0.81–1.60)	0.460	504
LIHC	1.52 (1.24–1.85)	1.10 (0.92–1.20)	0.420	375	1.00 (0.88–1.20)	0.760	351
LUAD	21.40 (19.7–23.2)	0.96 (0.79–1.20)	0.690	455	0.97 (0.80–1.20)	0.800	451
LUSC	21.40 (19.7–23.2)	**1.20 (1.00–1.50)**	**0.040**	**365**	**1.20 (1.00–1.50)**	**0.049**	362
OV	1.04 (0.95–1.15)	1.00 (0.38–2.60)	1.000	10	170 (0.21–15000)	0.130	10
PAAD	2.30 (2.08–2.53)	0.93 (0.72–1.20)	0.600	184	1.30 (0.94–1.80)	0.120	181
PRAD	0.85 (0.77–0.95)	0.26 (0.05–1.20)	0.091	484	0.26 (0.05–1.40)	0.120	484
READ	1.25 (1.14–1.37)	0.86 (0.52–1.40)	0.570	94	1.00 (0.60–1.70)	0.990	85
STAD	2.00 (1.67–2.39)	**1.30 (1.00–1.60)**	**0.022**	**393**	**1.30 (1.10–1.70)**	**0.016**	382
THCA	0.52 (0.35–0.78)	0.57 (0.28–1.20)	0.120	502	0.49 (0.21–1.10)	0.099	500
UCEC	0.75 (0.58–0.95)	0.88 (0.68–1.10)	0.330	425	1.00 (0.79–1.30)	0.840	425
UCS	0.83 (0.65–1.04)	1.10 (0.74–1.60)	0.690	57	1.10 (0.75–1.60)	0.640	57

Abbreviations: body mass index (body mass index), confidence interval (CI), deoxyribonucleic acid methylation (DNAm), number of patients (N), hazards ratio (HR), odds ratio (OR) and tumour-node-metastasis (TNM)

Significant HR associations are shown in bold

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

For the full, pre-menopausal and post-menopausal DNAm BMI associated HR analyses, it was also found that DNAm BMI, age and late TNM stage were all significant predictors of survival for the full and post-menopausal BRCA groups. No variables were significant predictors of survival for the pre-menopausal BRCA group (Table 4). Furthermore, in the subsequent analyses, ovarian cancer was excluded due to low patient numbers.

Table 4

DNAm BMI exposure survival analysis for breast cancer groups. The DNAm BMI exposure associated HRs were calculated by multivariable Cox proportional hazards analysis with adjustment for age at diagnosis, TNM stage (where available) and DNAm smoking exposure score, for the full, pre-menopausal and post-menopausal breast cancer groups.

Variables	BRCA full			BRCA pre-menopausal			BRCA post-menopausal		
	HR	P-value	N	HR	P-value	N	HR	P-value	N
BMI[1]	**1.60 (1.25–2.00)**	**	**774**	1.84 (0.95–3.60)	0.069	168	**1.44 (1.06–2.00)**	**0.018**	**495**
Smoking[1]	0.96 (0.76–1.20)	0.741	774	0.84 (0.41–1.70)	0.634	168	0.92 (0.69–1.20)	0.589	495
Age	**1.04 (1.02–1.10)**	***	**774**	0.95 (0.86–1.00)	0.229	168	**1.05 (1.03–1.10)**	***	**495**
Stage I & II	1.00 (reference)		555	1.00 (reference)		112	1.00 (reference)		369
Stage III & IV	**2.77 (1.83–4.20)**	***	**208**	1.77 (0.52–6.00)	0.362	53	**2.98 (1.78–5.00)**	***	**123**
Abbreviations: body mass index (BMI), breast cancer (BRCA), deoxyribonucleic acid methylation (DNAm),									
number of patients (N), and hazards ratio (HR)									
[1] DNAm exposure z-score									
Significant HR associations are shown in bold									
* p < 0.01, ** p < 0.001, *** p < 0.0001									

For each exposure the relationship between DNAm exposure associated cancer survival and reported exposure associated cancer risk in the 23 cancer types are shown in Fig. 4A. The DNAm exposure associated HRs and the reported exposure associated ORs for cancer survival and risk respectively, was significantly associated for the alcohol exposure (p = 0.022), and not significantly associated for the BMI and smoking exposures (p = 0.548, p = 0.193 respectively). The cancer types that had significant DNAm exposure associated HRs for cancer survival are shown with their reported exposure associated ORs for cancer risk in Fig. 4B. For the DNAm exposures and cancers that were significantly associated with survival for; kidney (KIRP), esophageal (ESCA) and head and neck (HNSC) cancers for higher alcohol consumption, pancreatic (PAAD) and post-menopausal breast (BRCA) cancers for higher BMI, and stomach (STAD), kidney (KIRC), bladder (BLCA) and lung (LUSC) cancers for smoking exposures; their

corresponding reported exposures were also associated with cancer risk, usually in the same direction. While for the DNAm exposures and cancers that were significantly associated with cancer survival for; bladder (BLCA) and brain (LGG) cancers for higher alcohol consumption, bladder (BLCA) cancer for higher BMI, and B-cell lymphoma (DLBC) cancer for smoking exposures; their corresponding reported exposures were not associated with cancer risk. Interestingly, the reported smoking exposure increased the risk of developing kidney (KIRC) cancer, but DNAm smoking exposure appeared to be protective in terms of prognosis.

Discussion

In this study we have used existing prediction models for the alcohol, BMI and smoking lifestyle exposures based on DNAm signatures to predict the patient's exposures based on their tumour DNA samples. Previous work has developed and validated these DNAm exposure signatures in numerous tissue samples, predominantly blood sample DNA, but this study is the first to our knowledge, to use tumour DNA to predict the exposures of the individuals. We first show that the DNAm exposure signatures observed in tumour DNA are correlated with the signatures as predicted from matching adjacent normal tissues for the alcohol and BMI exposures. This is important to address the potential limitation that tumour DNA methylation profiles change dramatically compared with the normal tissue in which they occur. We have then used these predicted DNAm exposures to investigate how these exposures relate to overall survival in the cancer patients. We find that specific cancer types have strong associations between poorer survival and higher alcohol consumption (bladder (BLCA), brain (LGG), esophageal (ESCA), and head and neck (HNCS) cancers), higher BMI (bladder (BLCA), pancreatic (PAAD) and post-menopausal breast (BRCA) cancers), and smoking (B-cell lymphoma (DCLB), stomach (STAD), bladder (BLCA), and lung (LUSC) cancers). While kidney (KIRC) cancer unusually was found to have improved survival with higher alcohol consumption and smoking exposures. For alcohol consumption we found a positive association between HRs and ORs across all cancers, indicating that for cancers where alcohol consumption is a significant risk factor, it is also associated with poorer survival.

For the smoking exposure, we found the normal tissue and tumour tissue did not correlated strongly and were separated in the hierarchical clustering. We propose two possible explanations for this Firstly, unlike the other two exposures, smoking is known to induce many mutations in CpG sites directly which could impact on observed DNA methylation patterns in tumour compared with normal. Alternatively, it could be that methylation patterns in the tumours represented in this analysis are more divergent for the smoking related CpG sites compared with the other exposure CpG sites.

Many of the findings in our study are consistent with the existing literature. The hazardous role of high alcohol in patients with esophageal (ESCA) and head and neck (HNSC) cancers (8, 9), and high BMI in breast (BRCA), bladder (BLCA) and pancreatic (PAAD) cancers (12–14), and smoking in stomach (STAD), lung (LUSC) and B-cell lymphoma (DLBC) cancers (16, 18, 19) was supported by studies that were based on clinical or self-reported phenotypes. However, we did not find studies supporting our findings of the hazardous role of high alcohol in patients with bladder (BLCA) and brain (LGG) cancers, and these

represent novel findings. Furthermore some reported associations of lifestyle exposures with cancer prognosis were not supported by our study. This includes the poorer cancer prognosis associations between colorectal cancer (6) and high alcohol, and ovarian (11) and colorectal (15) cancers and higher BMI. This lack of replication of previous findings could be due to the different patient cohorts used in these studies, low statistical power for these tumour types, or could reflect an interesting biological difference in the way the exposures are measured. For example, BMI often used in reported datasets, is measured by patients current weight and height is typically a single measurement used as a proxy for the measurement that may fluctuate throughout life, while the DNAm BMI exposure measurement may reflect a longer-term history of high or low adiposity.

This study has many strengths. Firstly, the large sample size of the TCGA Pan-Cancer collection, allowed us to examine and compare the effect of the lifestyle-associated DNAm exposures in multiple cancer types and granted us sufficient statistical power in the survival analysis. The pre-standardized molecular data prevented any influence caused by batch-effects or other technical confounders. The usage of the revised version of the clinical endpoint data also increased the accuracy of the survival analysis.

However, this study is not without limitations. Firstly, we acknowledge up-front that the variability in DNA methylation profiles in tumour DNA may influence the accuracy of these exposure predictions. Nevertheless, this prediction model can represent the biologically measured exposure rather than the phenotype itself reported by individuals. In the case of smoking, it has been confirmed that hypomethylation associated with the *AHRR* and *CYP1B1* gene induced by cigarette smoking were found in both lung tissue, blood and other tissues in the body (32). Therefore, it is not unexpected that the exposures can also be detected in tumour DNA. This biologically measured exposure may represent a more accurate representation than what can be achieved with questionnaires that ask about historical alcohol consumption with considerable recall bias.

Additionally, the DNAm exposure prediction model we used to quantify the lifestyle exposure was developed from methylation data measured in blood samples. However, in the TCGA dataset, DNA methylation was measured in the target organ, with the majority been taken from the primary tumour tissue. Whether these organ tissues have a consistent DNA methylation profile with the blood in terms of CpG sites associated with lifestyle exposure remained unclear. Although one study has pointed out that tissue from alveoli has a similar epigenetic profile with blood-derived sample at CpG sites associated with smoking exposure (32). We are unable to ensure that this is the case for the remaining organ tissues and lifestyle exposures, due to the lack of blood-derived DNA methylation data in the TCGA dataset. We are also unable to account for potential disparities in exposure or methylation associated with variables such as ethnicity and recruitment centre that may be biased in some tumour types compared with others as this data was not available in this dataset. While we were able to adjust for treatment response in some tumour types, we were not able to do that for all, therefore this could be improved in future studies. Another limitation lies with the missing values in the TCGA's DNA methylation data which prevented us from investigating the complete set of exposure associated CpG sites. In the future, blood-derived DNA methylation data measured in cancer patients could be used to validate our study. The consequence of

not including these missing CpG sites in the DNAm exposure prediction models cannot be assessed without comparison to the complete methylation data. There may also be unobserved confounding factors that have remained unadjusted, as we only adjusted for the most relevant confounding factors in consideration of the reduced statistical power.

Conclusions

In summary, we presented the lifestyle exposure mediated cancer risk and the survival risk in multiple cancer types. We found that DNAm exposure signatures can be measured in tumour DNA and are associated with poorer cancer survival in many cancers due higher alcohol consumption, higher BMI and smoking exposures. Cancer types whose survival probability is affected by the predicted DNAm exposures are also likely to have reported exposure cancer risk in the same direction, with few exceptions. The cancers that originated in organs with direct contact to the exposure, also tends to have a positive association between the cancer survival and cancer risk.

Abbreviations

BLCA: Bladder Urothelial Carcinoma, BRCA: Breast Invasive Carcinoma, CESC: Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma, CHOL: Cholangiocarcinoma, COAD: Colon Adenocarcinoma, DLBC: Diffuse Large B-cell Lymphoma, ESCA: Esophageal Carcinoma; GBM: Glioblastoma Multiforme, HNSC: Head and Neck Squamous Cell Carcinoma, KICH: Kidney Chromophobe, KIRC: Kidney Renal Clear Cell Carcinoma, KIRP: Kidney Renal Papillary Cell Carcinoma, LGG: Brain Lower Grade Glioma, LIHC: Liver Hepatocellular Carcinoma, LUAD: Lung Adenocarcinoma, LUSC: Lung Squamous Cell Carcinoma, OV: Ovarian Serous Cystadenocarcinoma, PAAD: Pancreatic Adenocarcinoma, PRAD: Prostate Adenocarcinoma, READ: Rectum Adenocarcinoma, STAD: Stomach Adenocarcinoma, THCA: Thyroid Carcinoma, UCEC: Uterine Corpus Endometrial Carcinoma, UCS: Uterine Carcinosarcoma

Declarations

Ethics approval and consent to participate

Not applicable for this study. All participants in the TCGA were originally recruited with informed consent in line with the TCGA Ethical Policies. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/history/policies

Consent for Publications

Not Applicable.

Availability of data and materials

All data are publically available from The Cancer Genome Atlas. Data for this project was downloaded from https://xenabrowser.net/

Competing interests

The authors declare no competing interests

Funding

KT carried out this study in partial fulfilment of his Cancer Informatics Masters in Research degree at Imperial College London. This work was in part supported by Ovarian Cancer Action.

Authors' contributions

KT performed the analysis conducted in this study; JS and JF supervised the study; all authors contributed to drafting the manuscript and all authors approved the final version prior to submission.

Acknowledgements

The authors acknowledge infrastructure support from the Imperial Experimental Cancer Medicine Centre, Cancer Research UK Imperial Centre, the National Institute for Health Research Imperial Biomedical Research Center, and the Ovarian Cancer Action Research Centre. The authors would like to thank the study participants, study staff, the doctors, nurses, and other healthcare staff and data providers who have contributed to the TCGA study cohorts. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

References

1. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. Lancet Lond Engl. 2014 Aug 30;384(9945):755–65.
2. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D, et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ. 2007 Dec 1;335(7630):1134.
3. Sugawara Y, Tsuji I, Mizoue T, Inoue M, Sawada N, Matsuo K, et al. Cigarette smoking and cervical cancer risk: an evaluation based on a systematic review and meta-analysis among Japanese women. Jpn J Clin Oncol. 2019 Jan 1;49(1):77–86.
4. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. Br J Cancer. 2015 Feb 3;112(3):580–93.
5. Wenbin D, Zhuo C, Zhibing M, Chen Z, Ruifan Y, Jie J, et al. The effect of smoking on the risk of gallbladder cancer: a meta-analysis of observational studies. Eur J Gastroenterol Hepatol. 2013 Mar;25(3):373–9.

6. Walter V, Jansen L, Ulrich A, Roth W, Bläker H, Chang-Claude J, et al. Alcohol consumption and survival of colorectal cancer patients: a population-based study from Germany. Am J Clin Nutr. 2016;103(6):1497–506.
7. Jayasekara H, English DR, Hodge AM, Room R, Hopper JL, Milne RL, et al. Lifetime alcohol intake and pancreatic cancer incidence and survival: findings from the Melbourne Collaborative Cohort Study. Cancer Causes Control CCC. 2019 Apr;30(4):323–31.
8. Chen Y-P, Zhao B-C, Chen C, Lei X-X, Shen L-J, Chen G, et al. Alcohol drinking as an unfavorable prognostic factor for male patients with nasopharyngeal carcinoma. Sci Rep. 2016 Jan 18;6:19290.
9. Ma Q, Liu W, Jia R, Long H, Zhang L, Lin P, et al. Alcohol and survival in ESCC: prediagnosis alcohol consumption and postoperative survival in lymph node-negative esophageal carcinoma patients. Oncotarget. 2016 Jun 21;7(25):38857–63.
10. Crispo A, Grimaldi M, D'Aiuto M, Rinaldo M, Capasso I, Amore A, et al. BMI and breast cancer prognosis benefit: mammography screening reveals differences between normal weight and overweight women. Breast Edinb Scotl. 2015 Feb;24(1):86–9.
11. Yang L, Klint A, Lambe M, Bellocco R, Riman T, Bergfeldt K, et al. Predictors of ovarian cancer survival: a population-based prospective study in Sweden. Int J Cancer. 2008 Aug 1;123(3):672–9.
12. Lin Y, Wang Y, Wu Q, Jin H, Ma G, Liu H, et al. Association between obesity and bladder cancer recurrence: A meta-analysis. Clin Chim Acta Int J Clin Chem. 2018 May;480:41–6.
13. Chan DSM, Vieira AR, Aune D, Bandera EV, Greenwood DC, McTiernan A, et al. Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. Ann Oncol Off J Eur Soc Med Oncol. 2014 Oct;25(10):1901–14.
14. Kasenda B, Bass A, Koeberle D, Pestalozzi B, Borner M, Herrmann R, et al. Survival in overweight patients with advanced pancreatic carcinoma: a multicentre cohort study. BMC Cancer. 2014 Sep 29;14:728.
15. Jayasekara H, English DR, Haydon A, Hodge AM, Lynch BM, Rosty C, et al. Associations of alcohol intake, smoking, physical activity and obesity with survival following colorectal cancer diagnosis by stage, anatomic site and tumor molecular subtype. Int J Cancer. 2018 15;142(2):238–50.
16. Avci N, Hayar M, Altmisdortoglu O, Tanriverdi O, Deligonul A, Ordu C, et al. Smoking habits are an independent prognostic factor in patients with lung cancer. Clin Respir J. 2017 Sep;11(5):579–84.
17. Li HM, Azhati B, Rexiati M, Wang WG, Li XD, Liu Q, et al. Impact of smoking status and cumulative smoking exposure on tumor recurrence of non-muscle-invasive bladder cancer. Int Urol Nephrol. 2017 Jan;49(1):69–76.
18. Han MA, Kim Y-W, Choi IJ, Oh MG, Kim CG, Lee JY, et al. Association of smoking history with cancer recurrence and survival in stage III-IV male gastric cancer patients. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2013 Oct;22(10):1805–12.
19. Simard JF, Baecklund F, Chang ET, Baecklund E, Hjalgrim H, -Olov Adami H, et al. Lifestyle factors, autoimmune disease and family history in prognosis of non-hodgkin lymphoma overall and subtypes. Int J Cancer. 2013 Jun 1;132(11):2659–66.

20. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003 Mar;33 Suppl:245–54.
21. Flanagan JM, Brook MN, Orr N, Tomczyk K, Coulson P, Fletcher O, et al. Temporal stability and determinants of white blood cell DNA methylation in the breakthrough generations study. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2015 Jan;24(1):221–9.
22. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017 05;541(7635):81–6.
23. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016 Oct;9(5):436–47.
24. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai P-C, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2018;23(2):422–33.
25. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic prediction of complex traits and death. Genome Biol. 2018 Sep 27;19(1):136.
26. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013 Mar 1;22(5):843–51.
27. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung C-H, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer. 2017 Jan 1;140(1):50–61.
28. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun. 2015 Dec 15;6:10192.
29. Kresovich JK, Xu Z, O'Brien KM, Weinberg CR, Sandler DP, Taylor JA. Methylation-based biological age and breast cancer risk. J Natl Cancer Inst. 2019 Feb 22;
30. Ambatipudi S, Horvath S, Perrier F, Cuenin C, Hernandez-Vargas H, Le Calvez-Kelm F, et al. DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. Eur J Cancer Oxf Engl 1990. 2017;75:299–307.
31. Johansson A, Palli D, Masala G, Grioni S, Agnoli C, Tumino R, et al. Epigenome-wide association study for lifetime estrogen exposure identifies an epigenetic signature associated with breast cancer risk. Clin Epigenetics. 2019 Apr 30;11(1):66.
32. Stueve TR, Li W-Q, Shi J, Marconett CN, Zhang T, Yang C, et al. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. Hum Mol Genet. 2017 01;26(15):3014–27.
33. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, et al. Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. JAMA Oncol. 2015 Jul;1(4):476–85.

34. Chiaffarino F, Ricci E, Cipriani S, Chiantera V, Parazzini F. Cigarette smoking and risk of uterine myoma: systematic review and meta-analysis. Eur J Obstet Gynecol Reprod Biol. 2016 Feb;197:63–71.
35. Kim K-N, Hwang Y, Kim K, Lee KE, Park YJ, Choi JY, et al. Active and Passive Smoking, BRAFV600E Mutation Status, and the Risk of Papillary Thyroid Cancer: A Large-Scale Case-Control and Case-Only Study. Cancer Res Treat Off J Korean Cancer Assoc. 2019 Feb 20;
36. Perez-Cornago A, Key TJ, Allen NE, Fensom GK, Bradbury KE, Martin RM, et al. Prospective investigation of risk factors for prostate cancer in the UK Biobank cohort study. Br J Cancer. 2017 Nov 7;117(10):1562–71.
37. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018 05;173(2):400–416.e11.
38. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct;45(10):1113–20.

Figures

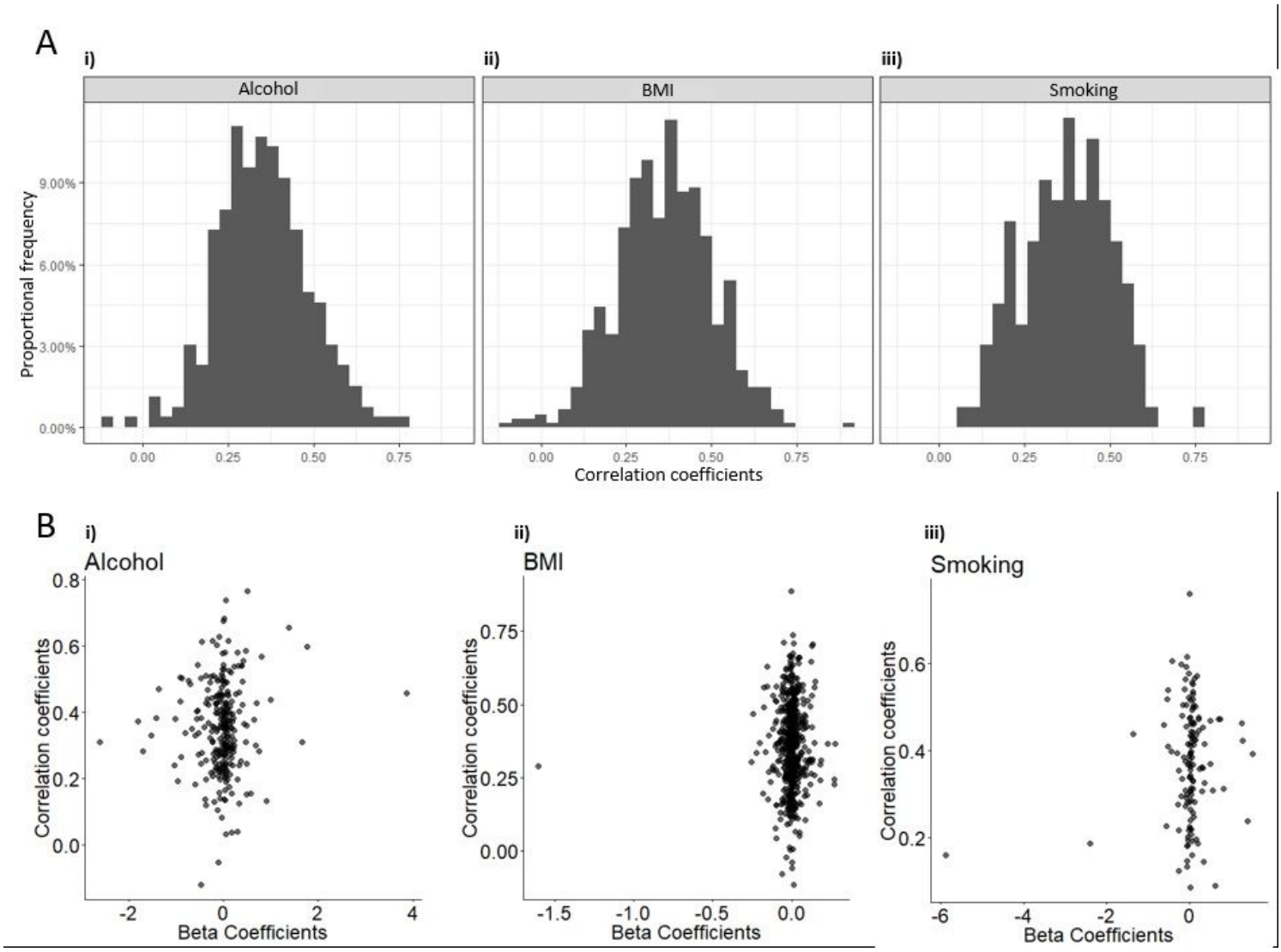

Figure 1

DNAm exposures and correlation coefficients between tumour and normal tissues. For each exposure, Spearman's rank correlation coefficients were calculated for each of the DNAm exposure signatures CpG sites beta-values between tumour and matched adjacent normal tissue for 696 patients. The correlation coefficients were then normalized into proportional frequencies by dividing by the number of total CpG sites associated with each exposure. The following were then plotted for each exposure. A) The i), ii), and iii) histogram plots represent the DNAm alcohol, BMI and smoking exposures respectively. Where patients tumour versus normal tissue correlation coefficients (X-axis) were plotted against their proportional frequencies (Y-axis). B) The i), ii), and iii) scatterplots represent the DNAm alcohol, BMI and smoking exposures respectively. In each plot the DNAm exposure signature CpG beta coefficients (X-axis) were plotted against corresponding patients tumour versus normal tissue correlation coefficients (Y-axis).

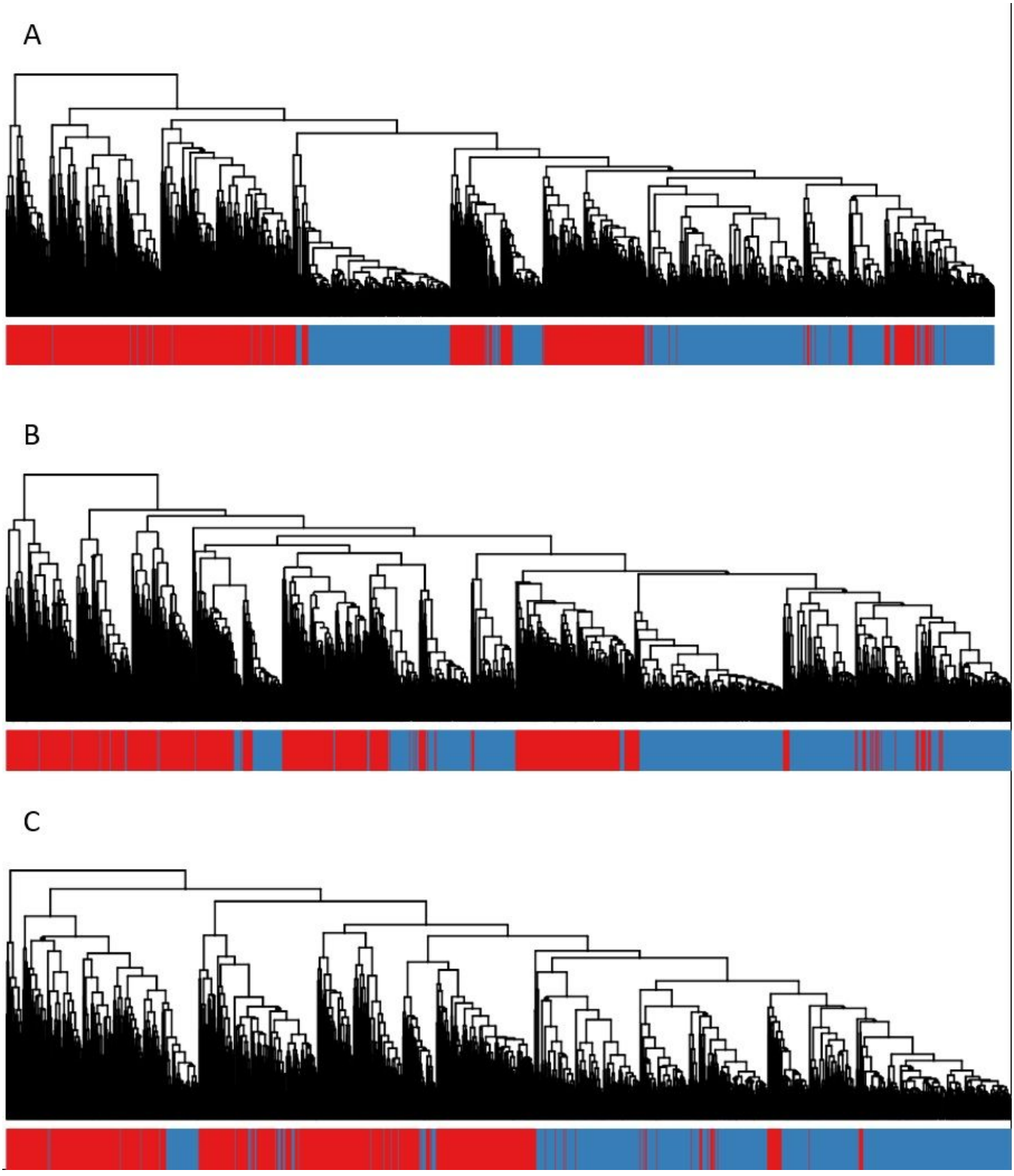

Figure 2

DNAm exposures and hierarchical clustering of tumour and normal tissues. For 696 patients their DNAm exposure CpG beta values from their tumour and matched adjacent normal tissues samples were hierarchically clustered by the Manhattan distance and visualised in dendrograms for the following DNAm exposures. A) DNAm alcohol exposure B) DNAm BMI exposure C) DNAm smoking exposure. Where the X-axis represents individual patient samples, with tumour tissue samples coloured red and

normal tissue samples coloured blue. BLCA, n= 21 ; BRCA, n= 75; CESC, n= 3; CHOL, n= 9; COAD, n= 38; ESCA, n= 16; GBM, n= 1; HNSC, n= 45; KIRC, n= 160; KIRP, n= 45; LIHC, n= 50; LUAD, n= 29; LUSC, n= 40; PAAD, n= 10; PCPG, n= 3; PRAD, n= 49 READ, n= 7; SARC, n= 4; STAD, n= 2; THCA, n= 54; THYM, n= 2; UCEC, n= 33.

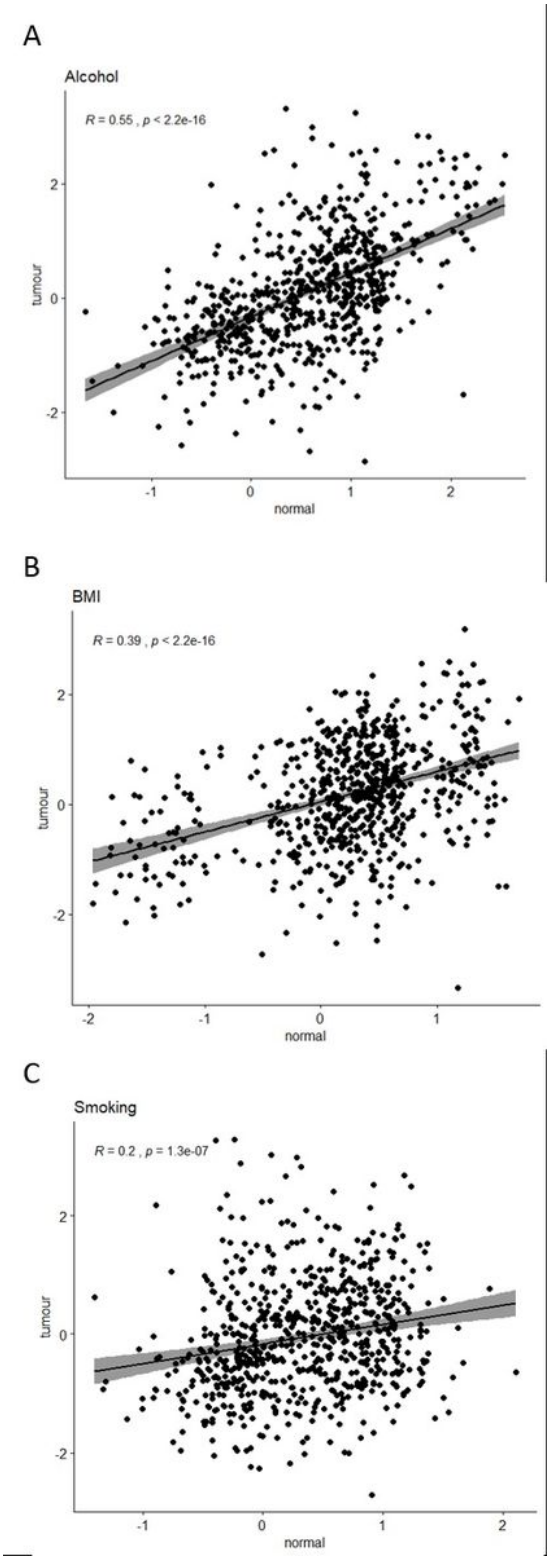

Figure 3

DNAm exposure z-scores and correlation between tumour and normal tissues. For 696 patients, their tumour and adjacent normal tissue DNAm exposure z-scores, were plotted in scatterplots and compared using Pearson's correlation for the following DNAm exposures A) DNAm alcohol exposure B) DNAm BMI exposure C) DNAm smoking exposure. Where the X-axis represents the DNAm exposure z-scores for patients normal tissue and Y-axis represents the DNAm exposure z-scores for patients tumour tissue .

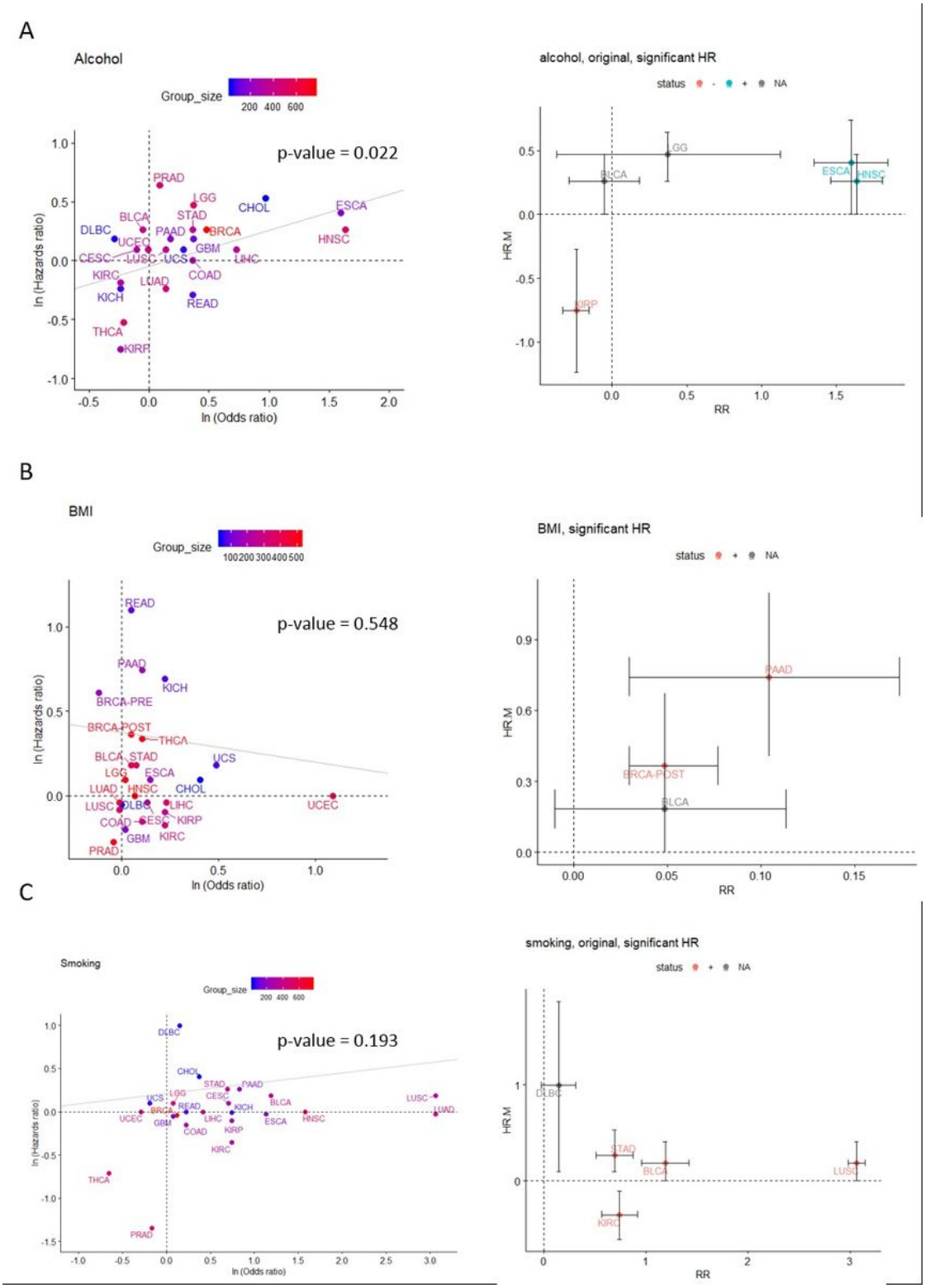

Figure 4

DNAm exposure cancer survivals versus reported exposure cancer risks. For the exposures and 23 cancer types, DNAm exposure associated HRs for cancer survival were calculated from multivariable Cox proportional hazard model analysis and their reported exposure associated ORs for cancer risk were gathered from the literature. The following were then plotted for each exposure, where i), ii), and iii) represent the alcohol, BMI and smoking exposures respectively. A) In each exposure plot the log-transformed ORs (X-axis) were plotted against their corresponding log-transformed HRs (Y-axis) for each cancer type. The cancer type sample size was indicated using a different colour, increasing in size from blue to red. The trend line denotes the log-transformed DNAm exposure HRs regressed on their respective log-transformed reported exposure ORs for each cancer type by linear regression, after adjustment for the group size of each cancer type. The vertical dashed black line represents log ORs of 0, and the horizontal dashed line represents log HRs of 0. B) Then of these, the cancer types that had significant HRs (Y-axis) were then plotted against their corresponding ORs (X-axis), for each exposure. The significance of the ORs was indicated using a different colour and symbol, where blue (+) represents significantly increased risk, red (-) represents significantly decreased risk and grey (NA) represents not significantly associated. The error bar represents the confidence interval for the ORs (vertical) and HRs (horizontal).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PanCanMethylationExposurePaperSupplementalTables110221.docx