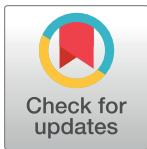


RESEARCH ARTICLE

Breast cancer survival and the expression of genes related to alcohol drinking

Hui G. Cheng¹, Agustin Gonzalez-Reymundez^{1,2}, Irene Li¹, Ania Pathak¹, Dorothy R. Pathak¹, Gustavo de los Campos^{1,2}, Ana Ines Vazquez^{1,2*}**1** Department of Epidemiology & Biostatistics, Michigan State University, MI, United States of America,**2** The Institute for Quantitative Health Science and Engineering, Michigan State University, MI, United States of America

* These authors contributed equally to this work.

* avazquez@epi.msu.edu

Abstract

Breast cancer is the leading cause of cancer-related disease in women. Cumulative evidence supports a causal role of alcohol intake and breast cancer incidence. In this study, we explore the change on expression of genes involved in the biological pathways through which alcohol has been hypothesized to impact breast cancer risk, to shed new insights on possible mechanisms affecting the survival of breast cancer patients. Here, we performed differential expression analysis at individual genes and gene set levels, respectively, across survival and breast cancer subtype data. Information about postdiagnosis breast cancer survival was obtained from 1977 Caucasian female participants in the Molecular Taxonomy of Breast Cancer International Consortium. Expression of 16 genes that have been linked in the literature to the hypothesized alcohol-breast cancer pathways, were examined. We found that the expression of 9 out of 16 genes under study were associated with cancer survival within the first 4 years of diagnosis. Results from gene set analysis confirmed a significant differential expression of these genes as a whole too. Although alcohol consumption is not analyzed, nor available for this dataset, we believe that further study on these genes could provide important information for clinical recommendations about potential impact of alcohol drinking on breast cancer survival.

OPEN ACCESS

Citation: Cheng HG, Gonzalez-Reymundez A, Li I, Pathak A, Pathak DR, de los Campos G, et al. (2020) Breast cancer survival and the expression of genes related to alcohol drinking. PLoS ONE 15 (2): e0228957. <https://doi.org/10.1371/journal.pone.0228957>

Editor: Qing-Xiang Amy Sang, Florida State University, UNITED STATES

Received: April 25, 2019

Accepted: January 27, 2020

Published: February 20, 2020

Copyright: © 2020 Cheng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The datasets analyzed during the current study are third-party data and can be requested from EGA. The Data Access Committee for METABRIC (<https://www.ebi.ac.uk/ega/dacs/EGAC00001000484>) controls 19 data sets. In this paper we have used normalized gene expression (study accession EGAS00000000083). Data has to be requested from METABRIC DAC (metabric@cruk.cam.ac.uk). The authors confirm they did not have any special access to this data.

Introduction

Breast cancer is the leading cause of cancer-related disease burden in women with an estimated 464,000 deaths worldwide [1]. Cumulative evidence supports a causal role of alcohol intake and breast cancer incidence [2–4]. Potential mechanisms include both global and site-specific changes induced by alcohol intake [5]. Nonetheless, evidence about alcohol consumption and breast cancer survival, an inherent element in cancer-related disease burden, has been mixed. Therefore, whether the observed pattern is a reflection of true causal relationship is unknown. For example, it is well possible that the observed relationships are confounded by self selection biases (e.g., “healthy drinker effect”) [6]. Given this context, it is of interest to examine the relationship between genes related to alcohol consumption and breast cancer

Funding: The study is supported by funds from the National Institute on Drug Abuse grant T32DA021129 (to HGC), as well as Michigan State University. The opinions expressed in this paper are solely those of the authors and do not reflect those of the sponsoring agencies, which included the National Institute of Drug Abuse, the National Institutes of Health, and Michigan State University.

Competing interests: The authors have declared that no competing interests exist.

survival, in order to shed new light on potential underlying mechanisms linking alcohol consumption prior to diagnosis and breast cancer survival.

Against this background, we seek to estimate the relationships linking the expression of genes related to alcohol drinking and breast cancer survival using a large database with modern quantitative genomics methods [7]. Our literature review identified the following groups of alcohol-related genes: genes involved in alcohol metabolism (*ALDH2*, *ADH1C*, *CYP2E1*, *AOX1*), DNA repair (*XDH*, *XRCC1*, *ERCC2*, *XPC*, *OGG1*), reactive oxygen species metabolism (*CBS*, *SOD2*, *GSTP1*), estrogen metabolism (*CYP1B1*, *COMT*), cell-to-cell contact (*ITGA5*) and folate metabolism (*MTHFR*) [8–11]. We first estimate the associations between the expression of each individual gene with breast cancer survival. Because sets of genes are often up- or down-regulated together, we conducted a gene set enrichment analysis (GSEA) to further elucidate whether the above mentioned genes were significantly up or downregulated in women with shorter or longer survival.

Materials and methods

Study population and sample

METABRIC: Patient data consisted of a sample of the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) data set [12,13] a 4-year retrospective cohort study with survival information from 1,977 breast cancer cases (1,415 luminal, 431 triple negatives, and 131 basal subtypes) of white Caucasian women between 21 and 96 years of age (median, 62 years of age). From this sample, a group of 1,674 patients received treatment in the form of radiotherapy, chemotherapy, and/or hormone therapy. Samples correspond to biopsies from primary tumor (breast) conserved as frozen tissue, and taken before patients underwent treatment. A total of 284 deaths occurred within four years after diagnosis (based on the time of maximum prediction accuracy of survival time for this data set [14], across different cancer subtypes: 131 luminal (40% deceased at fourth year), 431 triple negative (28%) and 131 Her2+ subtype (10%)). In this study, tumor grade was grade one (n = 950), grade two (n = 775), and grade three (n = 169), while tumor sizes ranged from 0.17 to 1.82 cm of diameter. Gene expression profiles from the tumor biopsies were normalized by the intensity of 49,473 microarray Illumina HT-12 v3 probes [12] and summarized at the gene level by the WGCNA R package [15]. Genes with more than 20% of missing data or zero variability were removed. The remaining missing data were replaced by the mean. The final number of genes after quality controls was 19,535. More detail about cohort and edition criteria can be found elsewhere [12,14]. The current study uses anonymized data accessed through Synapse (<https://www.synapse.org/Portal.html#!Synapse:syn1710250/wiki/27348>). Study was determined not “human subjects” by the Institutional Review Board (IRB) of Michigan State University (Study ID: STUDY00000996). Funding for METABRIC was provided by Cancer Research UK and the British Columbia Cancer Agency Branch. Normalized gene expression, clinical and survival data can be accessed from European Genome-phenome Archive (EGA, study accession [EGAS00000000083](https://ega-archive.org/studies/EGAS00000000083)) upon approval by METABRIC data access committee TCGA: The Cancer Genome Atlas ([16], n = 506) was used for validation. The dataset is younger than METABRIC and only a third of it has already 4 years of followup. In this dataset 437 patients were alive at 4 years. Covariables consisted of: age at diagnosis between 27 and 90 years old (median = 59 years), cancer subtype (0.18% triple negative, 78% luminal, 0.08% Her2+). Only 40% of patients in this dataset received some form of treatment. Gene expression count for the 18,322 genes was estimated from log RPKM counts from RNA-Seq obtained from Illumina HiSeq RNA V2 platform [17].

Statistical analysis

First, we performed a step-forward variable selection on a model including the median expression across all genes as response, using vital status, treatment, cancer subtype, tumor size, tumor grade and age as covariates. Next, we estimated the relationship between the expression of each gene in the data set and the selected covariates using the following statistical model followed by ANOVA:

$$y'_g = \mu_g + \alpha_g S + \beta_g C + \gamma_g A + \delta_g Sx C + \zeta_g Ax C + \eta_g Sx Ax C + \varepsilon_g \quad \text{Eq [1]}$$

where y'_g is the expression of the gene g (i.e. the expression level of the gene g across samples), μ_g is the general mean for the gene g , and S , A , C , $Ax C$, $Sx C$ and $Sx Ax C$ are survival status (S), age (A) cancer subtype (C), and the interactions between all terms (the Greek letters are the corresponding effects) (see [results](#)). The expression of gene g was considered associated with the survival of breast cancer when the estimated α_g was statistically significant based on p-values adjusted for multiple testing using the False Discovery Rate (FDR) method across all genes [18].

Next, to estimate the potential effect of groups of genes as aggregates, we performed a Gene Set Enrichment Analysis (GSEA) with the Limma R package [19]. The method starts by ranking the target genes based on their $\hat{\alpha}_g$ amongst the effects of the remaining genes. Then, a score is estimated, where the maximum can be considered as a modified version of the Kolmogorov-Smirnov (KS) statistics, weighted by the $\hat{\alpha}_g$ of each gene. The score obtained gives evidence to determine if the distribution of target genes differs or resembles the distribution of the remaining genes in the genome. A p-value was then obtained from a permutation test based on 10,000 iterations. Finally, if the target genes cluster on either the top or bottom of the list, the target genes would be considered as up- or down-regulated as a group.

Validation

Model described in Eq [1] was also adjusted in TCGA data. The p-values for survival are reported as validation of results.

Results

The significant covariates resulting from the variable selection procedure were vital status (S), cancer subtype (C), and age (A). These covariates and all their interactions were then tested with ANOVA in the models that have the expression of each gene as response. As shown in [Table 1](#), nine genes had differential expression between alive and death women with breast cancer from the 16 genes in the alcohol pathway. Three genes were up-regulated (i.e., *ITGA5*, *CBS*, and *SOD2*), while six were down-regulated (i.e., *XDH*, *XRCC1*, *MTHFR*, *CYP1B1*, *XPC*, and *GSTP1*) among individuals who died of breast cancer. Robust differences in the expressions of most of the genes under study were observed across the three cancer subtypes. [Fig 1](#) shows the results of the GSEA, in terms of the distribution of our target genes involved in alcohol consumption and breast cancer survival, against the expression level of the remaining genes in our data set. The plot shows that our target set of genes was significantly downregulated as a whole (p-value = 0.008), with most of its genes clustered at the bottom of the sorted vital status' effects ($\hat{\alpha}_g$). No statistically significant interactions of gene expression were found between survival, age and cancer subtypes.

[Fig 2](#) shows a visual representation of gene expressions per gene and by cancer subtypes. Cancer subtype based on our ANOVA results ([Table 1](#)) was significantly associated with the expression of all but 2 (*ADH1C* and *ERCC2*) alcohol-breast cancer risk related genes. The

Table 1. ANOVA results for the expression of genes related to alcohol metabolism. HUGO gene names and functions are presented along with gene expression effect on survival ($\hat{\alpha}_g$). The effect shows if the gene was up- or down-regulated in tumors from deceased patients at the fourth year. The p-values for a two-way ANOVA (FDR-adjusted across all genes) are also shown. ANOVA tests considered the effect of gene expression on vital status (S), cancer subtype (C), and patients age (A). Same models were validated on breast cancer data from The Cancer Genome Atlas (TCGA).

Genes	Function	$\hat{\alpha}_g$		Adjusted FDR p-values			TCGA
		METABRIC	TCGA	METABRIC			
				S	C	A	
<i>XRCC1</i>	DNA repair	-2.87	1.78	<1e-3	<1e-3	0.42	--
<i>CYP1B1</i>	Estrogen Metabolism	-2.31	1.19	<1e-3	<1e-3	<1e-3	--
<i>XPC</i>	DNA repair	-2.20	-0.16	<1e-3	<1e-3	0.01	0.03
<i>GSTP1</i>	ROS	-2.02	-0.18	<1e-3	<1e-3	<1e-3	0.01
<i>OGG1</i>	DNA repair	-1.95	-0.01	0.41	0.00	0.37	--
<i>MTHFR</i>	Folate Metabolism	-1.65	1.07	<1e-3	<1e-3	0.21	--
<i>ADH1C</i>	Alcohol Metabolism	-1.57	1.41	0.69	0.06	<1e-3	--
<i>AOX1</i>	Alcohol Metabolism	-0.93	-0.69	0.10	0.00	0.67	--
<i>XDH</i>	DNA repair	-0.13	-0.13	<1e-3	<1e-3	0.45	--
<i>ALDH2</i>	Alcohol Metabolism	0.29	-0.00	0.27	<1e-3	0.99	--
<i>ERCC2</i>	DNA repair	0.30	0.37	0.08	0.14	0.83	--
<i>COMT</i>	Estrogen Metabolism	0.50	0.07	0.73	<1e-3	0.51	--
<i>CBS</i>	ROS	0.93	0.26	<1e-3	<1e-3	0.48	--
<i>CYP2E1</i>	Estrogen Metabolism	1.28	-1.52	0.09	<1e-3	0.17	--
<i>SOD2</i>	ROS	1.41	-0.71	<1e-3	<1e-3	0.43	--
<i>ITGA5</i>	Cell-to-cell contact	1.64	0.37	<1e-3	<1e-3	0.27	0.05

<https://doi.org/10.1371/journal.pone.0228957.t001>

figure shows the direction of the expression for each subtype. The figure also suggests that the triple negative (TN) subtype was the group with largest inter-individual variation in gene expression, as compared to the other subtypes.

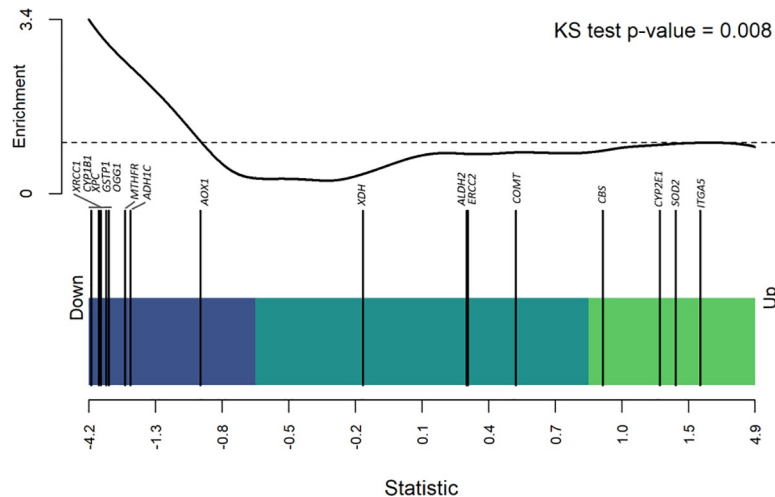


Fig 1. Gene-set enrichment analysis for gene expression across vital status. The x -axis corresponds with the sorted effects of vital status, estimated assuming the following probabilistic model $y'_g = \mu_g + \alpha_g S + \beta_g C + \gamma_g A + \delta_g Sx C + \zeta_g Ax C + \eta_g Sx Ax C + \epsilon_g$ for the expression of 19,535 genes. The light green and blue rectangles represent the up and down regulation regions defined by the third and first quartile, respectively. The vertical bars represent a target set of genes involved in alcohol metabolism. The black curve shows the estimated running enrichment score, and the dashed line shows the threshold for random disposition of the target set. The p-value was obtained by modified Kolmogorov-Smirnov (KS) test over the maximum value of the enrichment score. Genes in the target set: *XRCC1*, *CYP1B1*, *XPC*, *GSTP1*, *OGG1*, *MTHFR*, *ADH1C*, *AOX1*, *XDH*, *ALDH2*, *ERCC2*, *COMT*, *CBS*, *CYP2E1*, *SOD2*, and *ITGA5*.

<https://doi.org/10.1371/journal.pone.0228957.g001>

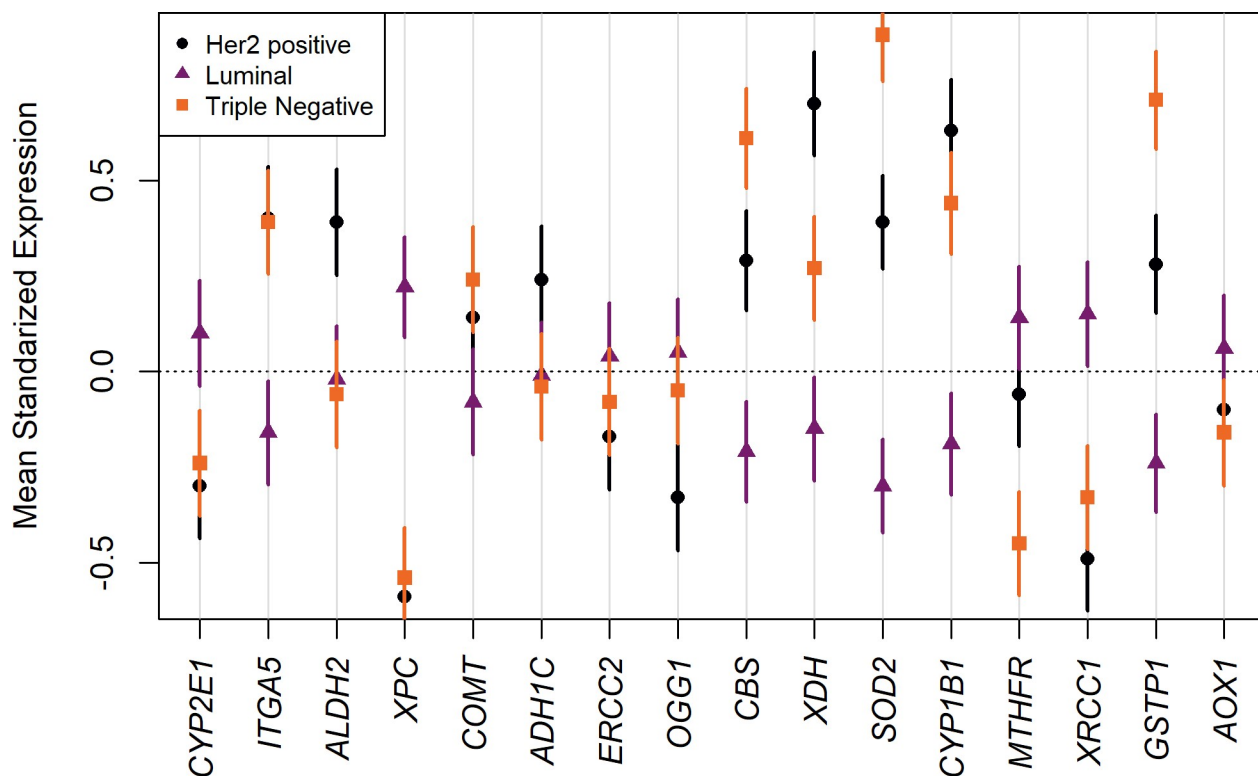


Fig 2. Target genes mean expression by cancer subtype. Bonferroni intervals were obtained considering the total number of pair-wise comparisons among all genes in the datasets. Intervals for alcohol related genes are plotted around the mean expression by cancer subtype.

<https://doi.org/10.1371/journal.pone.0228957.g002>

From the 9 genes significant in the Metabric dataset, genes *XPC*, *GSTP1*, and *ITGA5* were validated as statistically significant genes for survival in TCGA, and exhibiting the same direction of effects than in the Metabric dataset (Table 1). However, not all the significant genes in Metabric exhibited the same direction of effects in TCGA. The lack of validation of the other 6 genes, as well as incongruences in some of the effects directions, may indicate that either they are in fact not related to survival, or it could be simply lack of power due to the lower sample size of TCGA ($n = 506$) as compared to METABRIC ($n = 1,977$).

Discussion

In this study, we found that the expression of 9 out of 16 alcohol-related genes at the primary tumor were associated with breast cancer four-year survival, after correcting by cancer subtype and age, and 3 of these genes actually validate in a much smaller dataset. Four-year survival is a time point where maximum prediction accuracy was obtained with gene expression from breast cancer patients [14]. No evidence was found of interaction between the expression in different cancer-subtype, age and vital status.

It is known that alcohol consumption can increase the concentration of ROS and acetaldehyde, factors that can damage DNA and generate cancer-inducing mutations. Downregulation of *GSTP1* (gene involved in ROS metabolism) can elevate the concentration of ROS. This, together with downregulation of genes involved in DNA repair (e.g., *XDH*, and *XRCC1*) might enhance the noxious effects of oxidative agents. Decreased expression of *CBS* and *MTHFR* has been associated with increased levels of homocysteine in breast, which might also increase the risk of cancer [20]. However, *CBS* has been also reported as having oncogenic properties in

breast[21], as well as in colon and ovarian cancers[22]. Higher levels of *CYP2E1* have been linked with the production of procarcinogenics [23], while upregulation of *ITGA5* is known to be involved in higher invasiveness of tumor cells [24]. From these group of genes, *XPC*, *GSTP1*, and *ITGA5* were validated on breast cancer samples from TCGA. These genes may provide potential target for future studies on the mechanisms linking alcohol consumption, breast cancer risk and subsequent survival post diagnosis

The downregulation of *CYP1B1* in patients of shorter survival was unexpected. This gene is involved in the activation of pro-carcinogens, so we would expect the opposite direction in their expression. Here, 14 out of the 16 genes were differentially expressed across cancer subtypes. Differences were mostly between triple-negative and other subtypes. Due to the non-significant interactions between gene expression across vital statuses and cancer subtype (Table 1), this pattern might be explained by the different molecular heterogeneity exhibited by each subtype, with triple negatives being notoriously heterogeneous [25].

Implications of this study should be interpreted with the following limitations. This study does not include alcohol consumption level, thus gene expression level of alcohol-related genes may be up or down regulated due to either alcohol consumption or other factors. Finally, alcohol can induce a wide range of reactions after ingestion. Our list of genes is not exhaustive of all genes involved in these reactions. Future studies with an unbiased approach may provide more complete pictures for the alcohol-breast cancer risk -survival relationship.

In this study, we provide strong evidence linking the expression of 9 genes that are listed as involved in alcohol metabolism and breast cancer survival (*XRCC1*, *CYP1B1*, *XPC*, *GSTP1*, *MTHFR*, *XDH*, *CBS*, *SOD2*, and *ITGA5*) with FDR adjusted p-values lower than $1e-3$. Three of these genes are validated in an independent dataset. Future studies associating alcohol consumption before and after breast cancer diagnosis will be able to connect alcohol consumption and level of consumption with gene expression values in breast tissue, providing information for clinical recommendations about alcohol consumption among breast cancer survivors.”

Acknowledgments

The authors would like to thank the support from the National Institute on Drug Abuse grant T32DA021129 (to HGC), as well as Michigan State University. The opinions expressed in this paper are solely those of the authors and do not reflect those of the sponsoring agencies, which included the National Institute of Drug Abuse, the National Institutes of Health, and Michigan State University.

Author Contributions

Conceptualization: Hui G. Cheng, Agustin Gonzalez-Reymundez, Ana Ines Vazquez.

Data curation: Hui G. Cheng, Agustin Gonzalez-Reymundez, Ana Ines Vazquez.

Formal analysis: Hui G. Cheng, Agustin Gonzalez-Reymundez, Ana Ines Vazquez.

Funding acquisition: Hui G. Cheng.

Methodology: Hui G. Cheng, Agustin Gonzalez-Reymundez.

Supervision: Ana Ines Vazquez.

Writing – original draft: Irene Li, Ania Pathak, Dorothy R. Pathak, Gustavo de los Campos.

References

1. Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, et al. The Global Burden of Cancer 2013. *JAMA Oncol.* 2015; 1: 505. <https://doi.org/10.1001/jamaoncol.2015.0735> PMID: 26181261
2. Scoccianti C, Lauby-Secretan B, Bello P-Y, Chajes V, Romieu I. Female Breast Cancer and Alcohol Consumption: A Review of the Literature. *Am J Prev Med.* 2014; 46: S16–S25. <https://doi.org/10.1016/j.amepre.2013.10.031> PMID: 24512927
3. Park SY, Kolonel LN, Lim U, White KK, Henderson BE, Wilkens LR. Alcohol consumption and breast cancer risk among women from five ethnic groups with light to moderate intakes: The Multiethnic Cohort Study. *Int J Cancer.* 2014; 134: 1504–1510. <https://doi.org/10.1002/ijc.28476> PMID: 24037751
4. Shield KD, Soerjomataram I, Rehm J. Alcohol Use and Breast Cancer: A Critical Review. *Alcohol Clin Exp Res.* 2016; 40: 1166–1181. <https://doi.org/10.1111/acer.13071> PMID: 27130687
5. Wang J, Heng YJ, Eliassen AH, Tamimi RM, Hazra A, Carey VJ, et al. Alcohol consumption and breast tumor gene expression. *Breast Cancer Res.* 2017; 19: 108. <https://doi.org/10.1186/s13058-017-0901-y> PMID: 28899409
6. Mudry J. The poison is in the dose: The French Paradox, the healthy drinker and the medicalization of virtue. *Food, Cult Soc.* 2010; <https://doi.org/10.2752/175174410X12549021368144>
7. Margolin AA, Bilal E, Huang E, Norman TC, Ottestad L, Mecham BH, et al. Systematic Analysis of Challenge-Driven Improvements in Molecular Prognostic Models for Breast Cancer. *Sci Transl Med.* 2013; 5: 181re1–181re1. <https://doi.org/10.1126/scitranslmed.3006112> PMID: 23596205
8. Wang Y, Xu M, Ke Z-J, Luo J. Cellular and molecular mechanisms underlying alcohol-induced aggressiveness of breast cancer. *Pharmacol Res.* 2017; 115: 299–308. <https://doi.org/10.1016/j.phrs.2016.12.005> PMID: 27939360
9. Boffetta P, Hashibe M. Alcohol and cancer. *Lancet Oncol.* 2006; 7: 149–156. [https://doi.org/10.1016/S1470-2045\(06\)70577-0](https://doi.org/10.1016/S1470-2045(06)70577-0) PMID: 16455479
10. Roberts MR, Shields PG, Ambrosone CB, Nie J, Marian C, Krishnan SS, et al. Single-nucleotide polymorphisms in DNA repair genes and association with breast cancer risk in the web study. *Carcinogenesis.* 2011; 32: 1223–1230. <https://doi.org/10.1093/carcin/bgr096> PMID: 21622940
11. Ali AMG, Schmidt MK, Bolla MK, Wang Q, Gago-Dominguez M, Esteban Castela J, et al. Alcohol consumption and survival after a breast cancer diagnosis: A literature-based meta-analysis and collaborative analysis of data for 29,239 cases. *Cancer Epidemiol Biomarkers Prev.* 2014; 23: 934–945. <https://doi.org/10.1158/1055-9965.EPI-13-0901> PMID: 24636975
12. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* Nature Publishing Group; 2012; 486: 346–52. <https://doi.org/10.1038/nature10983> PMID: 22522925
13. Margolin AA, Bilal E, Huang E, Norman TC, Ottestad L, Mecham BH, et al. Systematic Analysis of Challenge-Driven Improvements in Molecular Prognostic Models for Breast Cancer. *Sci Transl Med.* 2013; 5: 181re1–181re1. <https://doi.org/10.1126/scitranslmed.3006112> PMID: 23596205
14. González-Reymúndez A, de los Campos G, Gutiérrez L, Lunt SY, Vazquez AI. Prediction of years of life after diagnosis of breast cancer using omics and omic-by-treatment interactions. *Eur J Hum Genet.* 2017; <https://doi.org/10.1038/ejhg.2017.12> PMID: 28272536
15. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, et al. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics.* BioMed Central; 2011; 12: 322. <https://doi.org/10.1186/1471-2105-12-322> PMID: 21816037
16. Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013; 45: 1113–1120. <https://doi.org/10.1038/ng.2764> PMID: 24071849
17. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLOS Comput Biol.* Public Library of Science; 2017; 13: e1005457. <https://doi.org/10.1371/journal.pcbi.1005457> PMID: 28545146
18. Yoav Benjamini YH. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B.* [Royal Statistical Society, Wiley]; 1995; 57: 289–300. Available: <http://www.jstor.org/stable/2346101>
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; gkv007-. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
20. Lin J, Lee I-M, Song Y, Cook NR, Selhub J, Manson JE, et al. Plasma Homocysteine and Cysteine and Risk of Breast Cancer in Women. *Cancer Res.* 2010; 70: 2397–2405. <https://doi.org/10.1158/0008-5472.CAN-09-3648> PMID: 20197471

21. Sen S, Kawahara B, Gupta D, Tsai R, Khachatryan M, Roy-Chowdhuri S, et al. Role of cystathionine β -synthase in human breast Cancer. *Free Radic Biol Med*. 2015; 86: 228–238. <https://doi.org/10.1016/j.freeradbiomed.2015.05.024> PMID: 26051168
22. Zhu H, Blake S, Chan KT, Pearson RB, Kang J. Cystathionine β -Synthase in Physiology and Cancer. *Biomed Res Int*. Hindawi Limited; 2018; 2018: 3205125. <https://doi.org/10.1155/2018/3205125> PMID: 30050925
23. Trafalis DT, Panteli ES, Grivas A, Tsigris C, Karamanakis PN. CYP2E1 and risk of chemically mediated cancers. *Expert Opin Drug Metab Toxicol*. 2010; 6: 307–319. <https://doi.org/10.1517/17425250903540238> PMID: 20073996
24. Zheng W, Jiang C, Li R. Integrin and gene network analysis reveals that ITGA5 and ITGB1 are prognostic in non-small-cell lung cancer. *Onco Targets Ther*. Dove Press; 2016; 9: 2317–27. <https://doi.org/10.2147/OTT.S91796> PMID: 27143927
25. Polyak K. Heterogeneity in breast cancer. *J Clin Invest*. American Society for Clinical Investigation; 2011; 121: 3786–8. <https://doi.org/10.1172/JCI60534> PMID: 21965334