EPiC Series in Computing

Volume 92, 2023, Pages 102–111

Proceedings of International Conference on Bioinformatics and Computational Biology (BICOB-2023)

An Association Analysis of Breast Cancer with Carotenoids

Samuel Neumann[1] and Mohamad El-hajj[2]

[1] University of Alberta, Edmonton, AB, Canada sfneuman@ualberta.ca
[2] MacEwan University, Edmonton, AB, Canada elhajjm@macewan.ca

Abstract

The environment and the exposure individuals carry throughout their lifetime can garner diverse effects on their health. This paper discusses the application of association analysis, to determine relationships between carcinogenesis and the human exposome. Human exposome data from the World Health Organization was analyzed to determine associations between human exposure and breast cancer. The discovered associations outline specific factors that may be associated with the prevention or causation of breast cancer. We discovered an association between biomarkers in specific biospecimens and breast cancer. Xanthophylls, measured in two different biospecimens, were determined to be associated with American breast cancer patients. The associations discovered may be of use in future cancer studies. This research is particularly interesting because of xanthophylls' relationship to retinol, inhibiting oncogenesis. Providing support and data for such associations will encourage more research on the exposome's effect on breast cancer and other conditions.

1 Introduction

Breast cancer is the leading cause of cancer-related deaths in women. It comprises roughly 18% of all cancers for females, with approximately 14,000 women dying due to it each year [8]. Although the death rates vary between regions of the world, many regions which have historically had a low risk of breast cancer have recently seen increased cases of the disease. Many factors may influence carcinogenesis in humans. One factor is diet, which has been researched and related to the development of breast cancer [8]. In particular, retinoids may prevent the growth of carcinogenic cells in the human body [8]. These preventative effects of retinoids are attributed to their integral role in cellular differentiation, and growth [6]. Tetraterpenoids, precursors to retinoids, may especially be used in cancer prevention [6].

Our study is interested in determining useful association rules based on the relationship between the human exposome and cancers. We are interested in determining associations useful to future research or prognosis. Although many association rules in the data set were meaningless or ambiguous, we could find association rules that linked certain exposures to biomarkers and biospecimens in breast cancer patients. Using biomarkers, it is possible to determine some physiological conditions related to a specific disease [14]. Our research was able to determine biomarker associations in breast cancer patients.

1.1 Terminology

- *Biomarker*: A biomarker is a measurement variable associated with disease outcome. In other words, it is a biological indicator that can be observed from outside the patient and implies that the patient is in a specific state. Usually, some medical state [13].
- *Exposure*: a single environmental factor or lifestyle factor an individual has been exposed to at some point during their life [16].
- *Exposome*: the collection of environmental exposures and lifestyle factors an individual is afflicted with throughout their lifetime, beginning from the prenatal period [16]. The Exposome can be thought of as the genome's environmental equivalent and has been the topic of many research words aimed to relate the Exposome to health conditions [17]
- *Carotenoid*: a family of fat-soluble chemical pigments present in red, yellow, orange, and dark-green fruit and vegetable [18].
- *Xanthophyll*: a class of oxygen-containing carotenoid pigments that perform a variety of critical roles in light-harvesting antenna assembly and function [11].
- *Cancer diagnosis*: the distinguishing of malignant tumors from benign tumors [5].
- *Cancer prognosis*: the likelihood of recurrence of cancer in individuals [5].

1.2 Biomarkers

Recall that a biological marker, or biomarker, is an external indicator for a patient, implying that the patient is in a specific state. Usually, some medical state [13]. Biomarkers, therefore, are not limited to chemicals alone. Any phenomenon indicating a person's medical state may be a biomarker; such indicators include chemical, physical, functional, molecular or biological measurements [13].

Biological markers are important in research. They are the most objective yet reproducible indicators used in modern medicine. Biomarker use as the endpoint of clinical trials has become commonplace in contemporary research [13]. Although biological markers need not be chemical compounds, the biomarkers studied in this work are chemicals.

1.3 Breast Cancer and the Environment

Before continuing to the association analysis, it is first critical to recognize that environmental contributions are underlying factors for many, if not most, breast cancer cases. Breast cancer incidence has increased fourfold between 1970 and 2014 in the United States. Similar increasing trends have also been observed worldwide.

Individual genetics are slow to change or evolve. Since the increase in breast cancer has been observed over two generations, we do not expect this change to be attributable to genetic change. Rather, we must focus on breast cancer's exposome and Heritability. Heritability is a statistical coefficient used in genetics to identify the proportion of the variation in a trait due to genetic differences between individuals. Various genome-wide association studies have reported the Heritability of breast cancer within the range of 0.097-0.13, which supports that breast cancer is largely determined by non-genetic factors, specifically environmental factors or the exposome.

Another critical consideration supporting the large environmental factors precluding breast cancer cases is age. Only 7% of breast cancer cases are diagnosed before the age of 40. This

has been attributed to many environmental factors, which may increase and change with age as we face more exposure. Some examples are breast density, diet, radiation exposure, weight, level of physical activity etc. The American Cancer Society has estimated that only 5-10% of breast cancer cases are indeed hereditary. Overall, overwhelming evidence indicates that most breast cancer cases are largely due to environmental reasons [3].

2 Related Works

Many knowledge discovery methods, such as clustering, classification, and association rule analysis, have been conducted on molecular biological data. Scherzer et al. enacted cluster analysis to determine clusters of biomarkers before determining the physiological associations of each biomarker cluster [12]. Association rule analysis on molecular biological data gained popularity around 2013 [14]. Szalkai et al. outlined a method of association rule analysis to determine combinatorial biomarkers of Alzheimer's disease [14]. Through association rules, they were able to show that certain sets of biomarkers may be an indicator for dementia, and other sets of biomarkers may be indicators for a lower probability of developing dementia [14].

Prediction of breast cancer survivability presents itself as a challenging research topic [4]. An erroneous prediction is harmful as it may give individuals who have breast cancer a false sense of security and hope. Statistical models are becoming more common as predictors of cancer survivability [1]. Examples of statistical models of breast cancer survivability prediction include using decision trees, artificial neural networks, and logistic regression to attempt to predict breast cancer survivability [4]. Abdelghani et al. compared the usage of decision trees, naive Bayesian classifiers, and neural networks when predicting breast cancer survivability [1]. They found that decision trees outperformed all other methods of classification [1].

Other research has attempted to use machine learning algorithms to perform breast cancer diagnosis or prognosis. Diagnosis and prognosis present medical personnel with great challenges, and using statistical models to aid in these predictions has revolutionized this part of the medical field [5]. Gupta et al. provide a survey on popular techniques used to diagnose cancers and cancer relapses [5].

Association rules and neural networks have been combined to create a cancer diagnosis system that can predict cancer diagnosis with high accuracy and confidence [7]. Karabatak et al. used association rules to reduce the dimensionality of data input to a neural network for breast cancer diagnosis [7]. Feature extraction is key for an accurate prediction model, and even the best classifiers will perform poorly if features are not chosen appropriately [7]. Using association rules, feature extraction and dimensionality reduction can be performed to select features most highly associated with the predicted labels [7].

3 Data Set

In this work, we were interested in analyzing cancers and their relations to biomarkers. We required a specific amount of cancer-biomarker associations in the data set. Figures 1.A and 1.B graphically display the number of cancer-biomarker links in the data. As seen by these figures, breast cancer was the most frequently occurring cancer in the data set and contained the most associations with distinct xanthophylls in the data. Our association rules, outlined in the *Results* section (section 5), were generated on the breast cancer subset of the data and take into account xanthophyll associations with breast cancer.

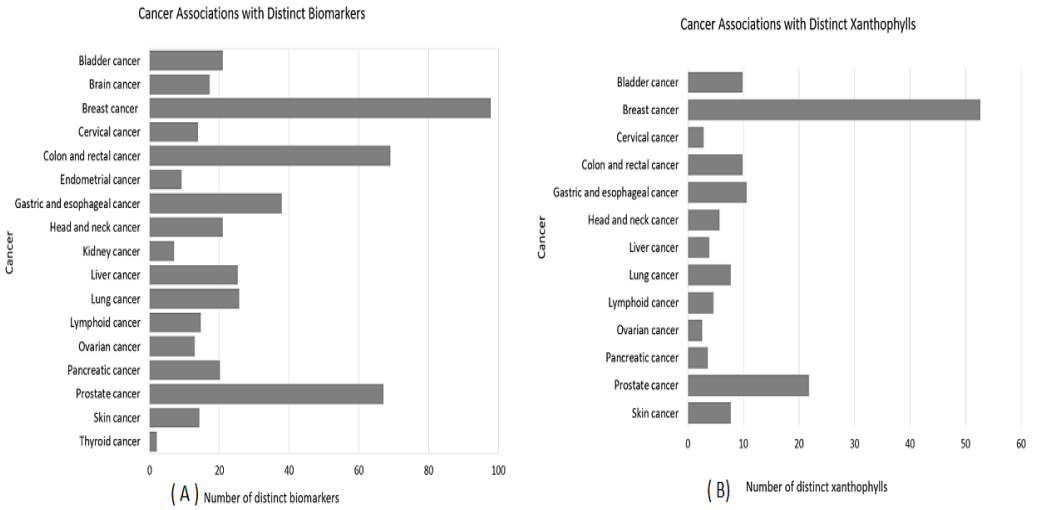

Figure 1: (A) Cancer-Biomarker Associations. Different cancers appeared in the data set. (B) Cancer-Xanthophyll Associations. Specific cancers appeared together with distinct xanthophyll biomarkers in the data set differing amounts of times.

3.1 Exposome-Explorer and Exposome-Explorer 2.0

We used the data set from Exposome-Explorer, [10]. This data was curated by the World Health Organization's International Agency for Research on Cancer. The data set described the human Exposome and was manually gathered and compiled from approximately 500 peer-reviewed studies. The Exposome-Explorer data contains various dietary and pollutant biomarkers measured in certain biospecimens and their correlation with exposures and diseases. The data contains over 8,000 associations between dietary intakes and biomarkers in associated biospecimens. The compilers of exposome explorer manually curated many peer-reviewed articles to capture the following information for each biomarker:

- *Publication*, together with the bibliographic details thereof.
- *Subject groups* which were studied in the publication, together with the population to which they belong.
- *Samples* taken from each subject group, together with the number of samples and time of biomarker sampling.
- *Biomarkers* and relevant information such as classification, molecular structure, atomic mass etc. Links to external chemical databases are included.
- *Biomarker measurements*, including the biospecimen from which the biomarker was measured from, any adjustments to concentrations of biomarkers
- *Correlations* between biomarkers and exposures or intakes, such as foods, food groups, or chemicals, together with the Pearson correlation coefficient and statistical significance.
- *Temporal reproducibility* as defined by the intraclass correlation coefficient. A high temporal reproducibility value is required for subject groups from which only a single

An update to the Exposome-Explorer dataset was published in 2019. The update expanded the number of biomarkers in the data set from 692 to 908; an additional 1,356 associations between dietary biomarkers and cancers were also included in the updated data set [9].

In addition to the information on each biomarker as listed above, the updated Exposome-Explorer incorporated the following data points into the data on dietary biomarkers:

- *Association of dietary biomarkers with food intakes.* This new data consist of a food intake which is associated with a specific dietary biomarker measured in specific biospecimens.

- *Associations of dietary biomarkers with cancer risk*, both significant and insignificant data was entered into the database.

For our study's purpose, it is important to note that the original Exposome-Explorer contained data on associations between chemical intakes (e.g. alpha-cryptoxanthin) and biomarkers. For the updated Exposome-Explorer, which linked associations between food intakes (e.g. coffee, citrus fruits) and dietary biomarkers, was added.

4 Methods

We began our study by selecting a few candidate data sets from the Exposome-Explorer data to perform association rule mining. We did not use every table in the data set, only those chosen to be relevant to our study. We thoroughly examined our data to determine the frequent data points in the candidate data sets. After determining the frequent data points, the data were binned into subsets. Each subset consisted of all data for each of the frequent data points. For example, we segmented the data into the subset of all data for which the biomarker was 25-hydroxycholecalciferol. For the frequent data sets, we considered past research on the topics. If much research had been conducted on the data formerly, the data set was rejected for association analysis. If the data had not been explored very thoroughly in past research, we attempted to mine for significant associations in the data.

4.1 Determination of Candidate Data Sets

To obtain meaningful association rules on biomarkers of exposure data, we were interested in using several sub-datasets from Expsome-Explorer. to perform effective association analysis to find associations between biomarkers of exposure cancer and intakes. Both food and chemical intakes were analyzed.

Our study aimed to determine associations between biomarkers and cancers to aid future research in determining biomarkers of interest for studies on specific cancers. As such, the sole incorporation of biomarkers and cancers was insufficient to ensure useful association rules. Multiple sub-datasets from Exposome-Explorer were consulted and studied, and we performed association analysis on the following data tables:

- *Biomarkers*, together with their classification, molecular data, and other attributes

- *Correlations with exposures*, which outlines chemical and food intakes/exposures and their associated biomarkers measured in specific biospecimens. The data set contains other useful attributes such as the country from which the studied samples were drawn, the correlation coefficient between exposure and biomarker, the studied cohort etc.

- *Correlations with cancer risks*, which models biomarkers found in specific biospecimens and cancer associated with each. The data set contains the size of each sample studied, the number of positive and negative instances in each sample, and other such attributes.

4.2 Selection of Subsets for Association Analysis

Although association rule mining was conducted on the entirety of the candidate data, the results were too general for our research purposes. This generality reduced the usefulness of the generated association rules. Therefore, we decided to look for association rules in specific subsets of the candidate data. This would cause the generated association rules to be much more specific. To determine a subset of the candidate data to mine, the candidate data set was converted to an SQL database using SQL Server and R. The candidate data set was then queried to discover how the data was laid out and which sets of items occurred most frequently together. This information would enable us to select a subset of the candidate data set upon which to mine for association rules. The subsets of data discussed below reveal data sets that were analyzed or data sets promising to find association rules but were not analyzed. This is not an exhaustive list of all subsets considered or mined upon.

We discovered that of the 17 cancers in the data set, the biomarker 25-hydroxycholecalciferol was associated with 16. This biomarker was measured in 8 different biospecimens in 12 different countries. Vitamin D was the only intake that was mapped to the biomarker 25-hydroxycholecalciferol in the data set. The biomarker 25-hydroxycholecalciferol produced the first subset of the data upon which we mined for association rules. Still, after performing an association analysis, the subset did not reveal any useful associations with cancer.

Retinol was the most frequently occurring class of biomarkers in the candidate data set. This class contained ten different biomarkers mapped to a total of 13 cancers in the data set. This class of biomarkers was measured in 9 different biospecimens across 12 countries. Since retinol is used in cellular differentiation, this subset of data was a promising data set to mine on. However, since past research has determined an association between retinol and cancers, we did not believe it productive to look for association rules between items that have already been deemed associated.

In our data set, breast cancer appeared more frequently than any other cancer, 452 times. This is over 177% more frequent than the next most frequently occurring cancer in the data set, prostate cancer. According to the data set used, breast cancer was linked to 97 different biomarkers. This subset revealed very promising data to find useful associations and was chosen as a data set upon which to perform association analysis. From this data set, we extracted useful association rules, which are outlined in the *Results* section (section 5) of this paper.

The United States was the most frequently occurring country in the data set. Association analysis was performed on the subset of data from the United States. The association rules generated from this subset of the data are outlined in the *Results* section (section 5).

4.3 Data Cleaning and Preparation

The first step in our process was to ensure that the data was clean and set in an appropriate format for association rule mining. We disregarded attributes that would have no meaning in our analysis. We then looked through all the attributes among all our relations to determine which attributes were missing excessive data points and could not be considered in our analysis. These attributes were considered incomplete and disregarded from our analysis.

To perform association analysis, we separated the data into a comma-delimited list. Next, we altered the data to be in *market basket form*. The market basket form breaks the data into

a delimited list of transactions, where each transaction contains all the items associated with that transaction. For example, the market basket form of one of our candidate data subsets had the form: *country*, *biomarker*, *biospecimen*, *intake*, *population*, and *cancer*.

4.4 Association Analysis

We performed association rule analysis using the apriori algorithm on each market basket form for each studied data subset [2]. Because association rule analysis is generally applied to large databases, the apriori algorithm is modelled to be very fast at determining these association rules [2].

We run the apriori algorithm with a minimum support of 0.05% and a minimum confidence of 30%. Recall that the support is the number of data points in our data set in which the variables in the association rule appear together. At the same time, confidence is the measurement of the confidence with which many variables are associated

5 Results

The results are lists of association rules generated from the data. We will begin with some background information that will increase the clarity of our results and how to read them. The results are separated into two tables, tables 2 and 3.

5.1 Reading Association Rules

Association rules are relationships that tell how a set of items is correlated with another set of items and outline the confidence with which these two items may be correlated. The rules are separated into a rule antecedent and a rule consequent. In the following sections, a rule antecedent is the list of items that appears before a $\Rightarrow$ symbol, and a rule consequent is the list of items that appears after the $\Rightarrow$ symbol. Association rules have support and confidence. The level of support is the percentage of data instances satisfying both the rule antecedent and the consequent. The confidence level of an association rule is the number of times the rule consequent appears in the data items that satisfy the rule antecedent.

5.2 Breast Cancer Subset Association Rules

The first set of results was found when mining the subset of data on breast cancer. The results are outlined in Figure 2. These association rules imply the following relationships:

- Measuring a xanthophyll biomarker in an American who has recently consumed fruits and vegetables is associated with finding that biomarker in the non-fasting plasma or serum.
- Taking a biosample from the non-fasting plasma or serum biospecimen of an American who has recently consumed fruits and vegetables is associated with finding a xanthophyll biomarker in that biosample.
- Measuring a xanthophyll biomarker in an American's non-fasting plasma or serum biospecimen is associated with the individual recently consuming fruits and vegetables.

The data implies that plasma and serum xanthophylls are fruit and vegetable intake biomarkers for American breast cancer patients. This means that after recently consuming fruits and vegetables, a xanthophyll biomarker is certain to be measured in the plasma or serum of an American breast cancer patient with the specified confidence.

Rule	Support	Confidence	Publications
(Fruits+Vegetables), (United States), (Xanthophylls) ⇒ (Plasma, non-fasting)	2.231%	59.18%	29
(Fruits+Vegetables), (Plasma, non-fasting), (United States) ⇒ (Xanthophylls)	2.231%	31.87%	29
(Plasma, non-fasting), (United States), (Xanthophylls) ⇒ (Fruits+Vegetables)	2.231%	34.94%	29
(Fruits+Vegetables), (United States), (Xanthophylls) ⇒ (Serum, unspecified)	15.38%	40.82%	20
(Fruits+Vegetables), (Serum, unspecified), (United States) ⇒ (Xanthophylls)	15.38%	40.00%	20
(Serum, unspecified), (United States), (Xanthophylls) ⇒ (Fruits+Vegetables)	15.38%	35.71%	20

Figure 2: Association rules from the breast cancer data subset.

Rule	Support	Confidence	Publications
(Breast Cancer), (Fruits+Vegetables), (Xanthophylls) ⇒ (Plasma, non-fasting)	13.54%	59.18%	29
(Breast Cancer), (Fruits+Vegetables), (Plasma, non-fasting) ⇒ (Xanthophylls)	13.54%	31.87%	29
(Breast Cancer), (Plasma, non-fasting), (Xanthophylls) ⇒ (Fruits+Vegetables)	13.54%	34.94%	29

Figure 3: Association rules from the USA data subset.

5.3 United States Subset Association Rules

The second set of results discovered in the data was found when mining the subset of data on the United States. The results are outlined in Figure 3. These association rules are relationships which imply the following:

- Upon eating fruits and vegetables and after being diagnosed with breast cancer, measuring a xanthophyll biomarker is associated with measuring that biomarker in the non-fasting plasma.
- Upon eating fruits and vegetables and after being diagnosed with breast cancer, taking a biomarker sample from the non-fasting plasma is associated with finding a xanthophyll biomarker in that biospecimen sample.
- Finding a xanthophyll biomarker in the non-fasting plasma of a breast cancer patient is associated with that patient consuming fruits and vegetables in the near past.

The data implies that plasma xanthophylls are fruit and vegetable intake biomarkers for breast cancer patients in the United States. These are very similar results to those discovered in the breast cancer subset of data. The main difference between the two results is a difference in perspective.

6 Discussion

The importance of our results comes from their relation to past and recent research. In this section, we comment on the interpretation of said results, briefly compare our result sets, and then discuss the relationship between our results and recent research. Our results can solidify past research as well as support current research attempting to determine the preventative nature of xanthophylls with cancer and determining the exposures for which xanthophylls may be biomarkers.

6.1 Comparison of Result Sets

As briefly mentioned, the main difference between the results in figures 2 and 3 is a difference in perspective. For the results on the breast cancer data subset, found in figure 2, the focus is

on breast cancer patients as a whole. These results imply that all breast cancer patients have the specified associations. For the results on the USA data subset, found in figure 3, the focus is narrowed down to only Americans. These results imply that Americans have the specified associations of all breast cancer patients. The second set of results discussed can be viewed as a verification of the first set.

6.2 Serum Xanthophylls as Biomarkers for Fruits and Vegetables

Previous work and research have suggested that serum xanthophylls may be biomarkers for fruit and vegetable intake [15]. This means that a xanthophyll in a person's serum indicates that the person has recently consumed fruits and vegetables. This research has only suggested this association between fruit and vegetable intake and xanthophylls. Our results indicate that serum and plasma xanthophylls are biomarkers for fruit and vegetable intakes. In this way, our results can contribute to and solidify the recent studies on the relationship of xanthophylls with these dietary exposures. Our results indicate that xanthophylls are biomarkers for fruit and vegetable intake in American breast cancer patients.

6.3 Summary

The discovered results indicate that, according to our data, there is a relationship between xanthophylls and breast cancer in the United States. In addition, our results outline a relationship between non-fasting plasma and serum xanthophylls as biomarkers for fruits and vegetables in American breast cancer patients.

Our results are interesting since they indicate an association between American breast cancer patients and the xanthophyll biomarkers. This correlation does not imply causation, but it may be a first step in identifying a relationship between breast cancer and xanthophylls. Note that our data imply that all these data points are associated but do not tell whether this is a positive or negative association. From our data, we cannot tell whether the association between breast cancer and xanthophylls is positive, that xanthophylls help to prevent breast cancer or breast cancer symptoms, or negative, that xanthophylls seem to cause or worsen breast cancer symptoms.

However, an interesting note is that, in past research, carotenes have been identified as possibly having a preventative effect on carcinogenesis [6]. Xanthophylls are related to carotenes. It may be possible that this relationship also hints at a protective effect of xanthophylls against breast cancer.

7 Conclusion

According to our study, xanthophylls may be correlated with breast cancer. Our results unveil an association between this biomarker and cancer that previous researchers could not find. In addition, recent research has suggested that high serum xanthophylls are biomarkers for fruit and vegetable intakes. Our discoveries have solidified this suggestion and support that high serum xanthophylls are biomarkers for fruit and vegetable intake in American breast cancer patients. These association rules may be used as guidelines for the early detection of breast cancer.

References

[1] Benyoucef Abdelghani and Erhan Guven. Predicting breast cancer survivability using data mining techniques. In *Second International Conference on Software Technology and Engineering*, 2006.

[2] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, page 487–499, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.

[3] Carey K. Anders, Rebecca Johnson, Jennifer Litton, Marianne Phillips, and Archie Bleyer. Breast cancer before age 40 years. *Seminars in Oncology*, 36(3):237–249, 2009.

[4] Dursun Delen, Glenn Walker, and Amit Kadam. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2):113 – 127, 2005.

[5] Shelly Gupta and Anand Sharma. Data mining classification techniques applied for breast cancer diagnosis and prognosis. *Indian Journal of Computer Science and Engineering*, 2, 04 2011.

[6] Charles H. Hennekens, Sherry L. Mayrent, and Walter Willett. Vitamin a, carotenoids, and retinoids. *Cancer*, 58(S8):1837–1841, 1986.

[7] Murat Karabatak and M. Cevdet Ince. An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*, 36:3465 – 3469, 2009.

[8] K McPherson, C M Steel, and J M Dixon. Abc of breast diseases. breast cancer-epidemiology, risk factors, and genetics. *BMJ (Clinical research ed.)*, 321:624–628, Sep 2000.

[9] et al Neveu. Exposome-Explorer 2.0: an update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Research*, 48(D1):D908–D912, 11 2019.

[10] Vanessa et al Neveu. Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Research*, 45(D1):D979–D984, 10 2016.

[11] Barry J. Pogson, Krishna K. Niyogi, Olle Björkman, and Dean DellaPenna. Altered xanthophyll compositions adversely affect chlorophyll accumulation and nonphotochemical quenching in arabidopsis mutants. *Proceedings of the National Academy of Sciences*, 95(22):13324–13329, 1998.

[12] Rebecca Scherzer, Sanjiv J. Shah, Eric Secemsky, Javed Butler, Carl Grunfeld, Michael G. Shlipak, and Priscilla Y. Hsue. Association of biomarker clusters with cardiac phenotypes and mortality in patients with hiv infection. *Circulation: Heart Failure*, 11(4):e004312, 2018.

[13] Kyle Strimbu and Jorge A. Tavel. What are biomarkers? *Current Opinion in HIV and AIDS*, 5(6):463–466, 11 2010.

[14] Balazs Szalkai, Vince K Grolmusz, and Vince I Grolmusz. Identifying combinatorial biomarkers by association rule mining in the camd alzheimer's database. *arXiv.org*, Dec 2013.

[15] Sara E Thomas and Elizabeth J. Xanthophylls. *Advances in Nutrition*, 9(2):160–162, 2018.

[16] Christopher Paul Wild. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology and Prevention Biomarkers*, 14(8):1847–1850, 2005.

[17] Christopher Paul Wild. The exposome: from concept to utility. *International Journal of Epidemiology*, 41(1):24–32, 01 2012.

[18] Xuehong et al Zhang. Carotenoid intakes and risk of breast cancer defined by estrogen receptor and progesterone receptor status: a pooled analysis of 18 prospective cohort studies. *The American Journal of Clinical Nutrition*, 95(3):713–725, 01 2012.