Food Chemicals in Epigenetic Targets: Towards an Epi Food Chemical Database

K. Eurídice Juárez-Mercado,[1] Juan F. Avellaneda-Tamayo,[1] Hassan Villegas-Quintero,[1] Ana L. Chávez-Hernández,[1] Claudia Daniela López-López,[2] José L. Medina-Franco[1,*]

[1] DIFACQUIM research group. Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México.

[2] PECEM, School of Medicine, Universidad Nacional Autónoma de México.

* Correspondence: medinajl@unam.mx; Tel. +52-55-5622-3899

Abstract

There is an increasing awareness of the importance of epigenetics to understand disease etiologies and develop novel therapeutics. Concomitantly, the renewed interest in epigenetic processes and their relationship with food chemicals has been reflected by an increasing number of publications in the past few years. However, there is a lack of a recent systematic and quantitative study that accounts for the most recent advances in the area associating the chemical structures of the food and natural product components with their biological activity. Here, we analyze recent advances and discuss the status of food chemicals and their intersection with natural products in epigenetic research. We discuss the most investigated diseases and potential therapeutic applications associated with food chemicals and natural compounds ingested in the diet. Using chemoinformatics tools, we compared quantitatively chemical contents, structural diversity, and coverage in the chemical space of food chemicals reported with epigenetic activity. As part of this work, we built and curated a compound database of food and natural product chemicals annotated with structural information, epigenetic target activity profile, and main source of the food chemical or natural product, among other relevant features. The compounds are cross-linked with identifiers from other major public databases such as FooDB and the COlleCtion of Open Natural ProdUcTs, COCONUT. The compound database is freely accessible at https://github.com/EuridiceJuarez/EpiFoodChemicalDatabase/blob/main/EpiFoodChemicalDatabase.csv.

Keywords: chemical space; databases; epigenetics; food chemicals; foodinformatics; natural products; structure-activity relationships.

1. Introduction

The concept of epigenetics has changed since it was first introduced in the 1940s by Conrad Waddington to describe "the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being"[1] Nowadays the meaning of epigenetics is widely accepted as the study of the heritable changes in the gene expression profile that do not entail a change in DNA sequence but modifies on the accessibility of the code via DNA methylation, modifications of amino acids on the amino-terminal tail of histones and non-coding RNAs.[1–3] It has been proposed that these changes could be classified into three types: direct epigenetics, which occur in the lifespan of a person, within indirect epigenetics define the ones that occur inside the womb, due to events during gestation and across indirect epigenetics, referred to those changes that affected the individual predecessors and somehow, maybe through changes in the gametes or intrauterine environment setting, are transmitted across generations.[2] The immense interest shown in the field lead to the development of many studies showing the link between epigenetic changes and certain diseases such as diabetes, heart failure, cancer, inflammatory bowel diseases, and neurodegenerative diseases, among others.[4–7]

Certain enzymes have been described as having a key role in these epigenetic modifications: DNA methyltransferases (DNMTs), in charge of the covalent addition of a methyl group to the DNA leading to the repression of certain genes; histone acetyltransferases (HATs) with the function of the acetylation of histone proteins, allowing the chromatin structure to open and become more transcriptionally active,[8] and histone deacetylases (HDACs), which regulate the deacetylation of histones, leading to a hypoacetylation towards heterochromatin and gene suppression.[9] Thus, the search for molecules that could hit these targets began, and the term "epidrugs" was coined to describe chemical compounds that alter DNA and chromatin structure, promoting the disruption of transcriptional and post-transcriptional modifications by the inhibition of DNMTs and HDACs, mainly. As of 2022, several compounds have been approved by the Food and Drug Administration of the USA for clinical use while other compounds are chemical probes. Examples of approved drugs are azacytidine (DNMT1 inhibitor), 5-aza-2'deoxycytidine (DNMTs and HDACs inhibitor), procaine (DNMTs inhibitor), hydralazine (DNMTs inhibitor), vorinostat (HDACs inhibitor), romidepsin (HDACs inhibitor), panobinostat (HDACs inhibitor), and belinostat (HDACs inhibitor).

Nanaomycin A is a promising probe molecule (DNMT3b inhibitor).[10-14] The chemical structures are shown in Figure 1.

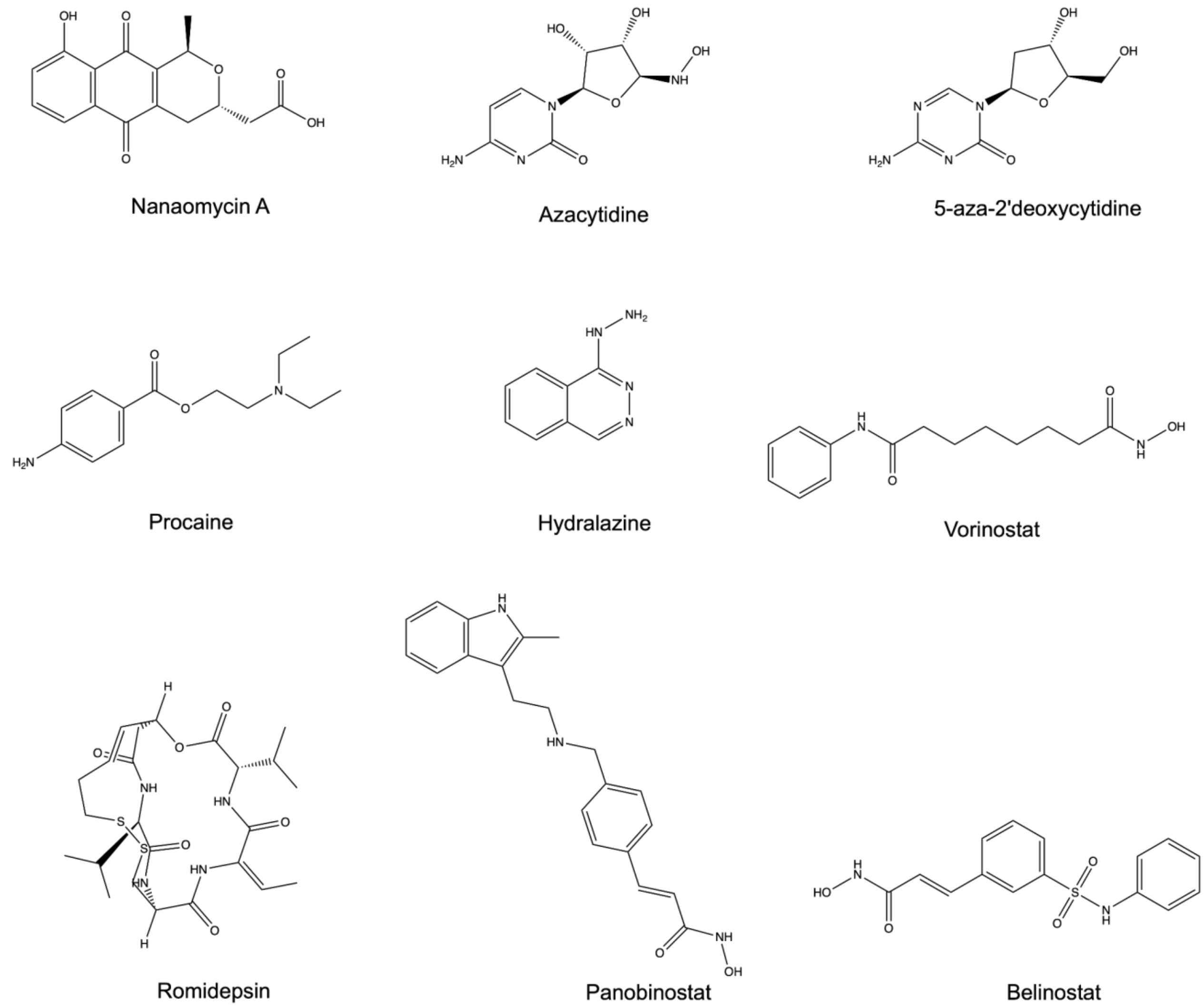

Figure 1. Chemical structures of representative drugs and drug candidates.

One of the most promising areas of this search is the field of nutriepigenomics, focused on the study of the interaction between food nutrients and genome through epigenetic mechanisms, modulating the overexpression or silencing of specific genes and metabolic responses.[15–17] The interaction between nutrition, epigenetic targets, and the development of certain diseases such as type I and type II diabetes, inflammatory diseases, liver fibrosis, and cancer have been discussed in the last few years, leading to new alternatives to mitigate the damage or prevent such conditions.[4,5,9,15,18–21]

Using chemoinformatics to analyze natural products[22] and food chemical data sets is increasingly widespread. The term foodinformatics, coined in 2014,[23] captures the application of chemical information to food science. Several works focused on the contents and diversity of food chemicals have been published, yielding useful information to organize and mine chemical information associated with food chemicals, which, ultimately, is at the core of informatics applications in chemistry.[24] Similarly, chemoinformatics has a growing interest in natural product research,[25] giving rise to the sub-field of natural products informatics.[26] Notable examples of the applications of chemoinformatics to food chemistry and natural product research are the development of large compound databases such as FooDB[27] and the Collection of Open Natural Products (COCONUT).[28] Despite the increasing evidence of the effect of food and natural products chemicals on epigenetic targets, to the best of our knowledge, there is not a comprehensive survey of the effect of food molecules on different epigenetic targets, rather than focusing on a specific disease or a specific epigenetic target family.

The goal of this study was to analyze the recent progress of research on food chemicals and food components acting with epigenetic targets, building a compound database that integrates the information on the chemical structure of food chemicals and other natural products with the epigenetic activity profile. The scientific papers and compound database were analyzed using chemoinformatics, data mining, and visualization approaches to identify: 1) the most frequent epigenetic targets and related therapeutic areas associated with food chemicals reported so far; 2) the food chemicals and other natural products most studied along with their epigenetic activity profile. The chemical structure contents, diversity, and coverage in the chemical space of the compounds in the molecular database were evaluated using quantitative methods and data visualization techniques. Since the chemical space of a compound data set depends on the structure representation, we characterize the recently termed "chemical multiverse," e.g., chemical space generated with multiple structure representations.[29] As part of the analysis, we explored the relationships between the chemical structures and the epigenetic activity profile using the approach of structure-property landscapes.[30]

2. Methods

2.1. Literature search and analysis

We conducted a meta-analysis of the literature of research papers published between 2017-March 2023 in peer-reviewed journals with digital object identifier (DOI) number, documenting the research of food chemicals interacting with epigenetic targets with potential therapeutic applications or disease prevention. The literature search was done in PubMed[31] and Web of Science Core Collection[32] databases using the following search terms: (("epigenetics" AND "food chemical(s)") OR *("epigenetics" AND "natural products")* OR ("epigenetics" AND "therapeutic application") OR ("epigenetics" AND "disease") OR ("epigenetics" AND "drug discovery") OR ("epigenetics" AND "drug development") OR "epigenetic targets" OR "epigenetic therapy" OR "epigenetic mechanisms" OR "epigenetic regulation" OR "epigenetic modifiers" OR "epidrugs" OR "nutritional epigenetics" OR "nutrigenetics"). As part of the analysis, the dietary compounds were determined in the abstract of the selected papers. Then the most common therapeutic indications associated with these compounds were selected in the related papers. Additional analyses were performed after assembling and annotating a compound database described in Section 2.2.

2.2. Compound database of food and natural product chemicals annotated with epigenetic activity

Based on the literature search and analysis described in Section 2.1, a compound database herein termed "Epi Food Chemical Database" was assembled using Google Sheets. The chemical structures were represented using the linear notation Simplified molecular-input line-entry system (SMILES).[33] The compound database was annotated with the following information: compound name; the International Chemical Identifier (InChI); the hashed version of InChI (InChIKey); main food source; if available, link of the compound to the FooDB or COCONUT databases (using the corresponding identifiers in those public databases); reference to the peer-reviewed paper using the DOI number; activity profile with the epigenetic targets for which the given compound has reported activity. To facilitate subsequent analysis and rapidly identify trends in the data, the activity profile was represented as a vector of "1"s and "0"s to indicate if the compound has or not reported activity with a given epigenetic target, respectively.

2.3. Chemoinformatic analysis of the chemical database

The content and diversity of the chemical structures of the 187 compounds in the Epi Food Chemical Database was analyzed under three main type of analysis: a) scaffold content and chemical diversity using structural fingerprints and chemical scaffolds; b) distribution in chemical space, and c) descriptive structure-activity relationships based on the concept of activity, or more general, property landscapes.[34] Each of the three types of analysis is described below.

2.3.1. Chemical content and diversity analysis

The scaffold content analysis was based on the definition of Bemis and Murcko,[35] which considers a scaffold as the rings in a molecule and the connectors of them, the analysis was performed using an in-house code in Python with the modules MurckoScaffold of RDKit library. Also, the chemical structures of the compound database were analyzed using well-established protocols and broadly used to characterize or assess the chemical diversity, namely, scaffold contents, and structural diversity using four molecular fingerprints: Molecular ACCEs System (MACCS) Keys (166-bits); Extended Connectivity Fingerprints (ECFP) radius 2 and 3; and RDKit fingerprints. The similarity analysis was calculated using the Jaccard-Tanimoto index.[36]

2.3.2. Visualization of the chemical space

To visualize the chemical space of the compounds in the Epi Food Chemical Database, we generated a t-distributed stochastic neighbor embedding (t-SNE). This technique involves nonlinearly reducing dimensions by creating Gaussian probability distributions across high-dimensional space and then utilizing them to enhance a Student t-distribution within a lower-dimensional space through optimization. The lower dimensional space conserves pairwise similarities from the original higher dimensional space, resulting in clustering within the embedding space without a notable loss of the structural information.[37-38] In the present work, a t-SNE was performed for the Epi Food Chemical Database and FooDB.

2.3.3. Structure-epigenetic activity profile

For the 187 compounds in the Epi Food Chemical Database, we computed all pairwise fingerprints based on the structural similarity of the chemical structures and the pairwise epigenetic activity profile similarity using the Jaccard-Tanimoto coefficient in both cases. The fingerprint-based similarity was calculated with four different fingerprints: ECFP4, ECFP6, MACCSKeys and RDKit fingerprints.[39] In total, 17,590 pairwise comparisons were computed for each fingerprint (including self-comparisons) and 17,430 pairwise comparisons for each fingerprint (excluding self-comparisons). The structure vs. epigenetic activity profile similarity was plotted in a scatter plot reminiscent of the structure-activity similarity (SAS) maps.[40-43] Figure 2 shows a prototype plot of a SAS map where the epigenetic activity profile similarity is plotted on the Y-axis while the fingerprint-based structural similarity is plotted on the X-axis. A SAS map can be roughly divided into four regions as described in Figure 2; in Region I are pair of compounds with very similar activity profiles but very different structural similarity, in Region II are pair of compounds with high structural similarity as similar activity profiles. Region III identifies pairs of compounds with high structural similarity but very different activity profiles. In Region IV are pairs of compounds with very similar activity profiles but very different structure similarity, that is activity or property cliffs.

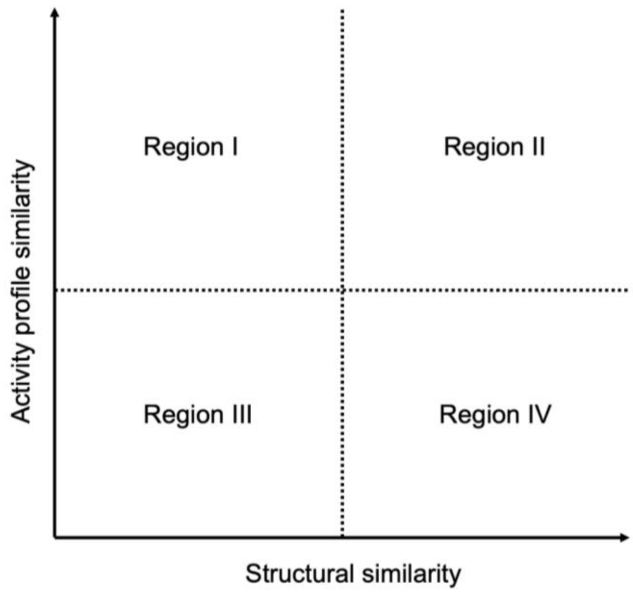

Figure 2. Prototype plot of a structure-activity similarity (SAS) map. Pairs of compounds in regions I and III have low structural similarity, while those in regions II and IV have high structural similarity. Pairs of compounds in regions I and II have a high similarity in their epigenetic activity profiles, although the chemical compounds in regions III and IV hold very different epigenetic activity profiles.

3. Results and discussion

3.1. Literature analysis

The literature search revealed that the number of peer-reviewed papers found in PubMed and Web of Science using the search terms described in the Methods was 7,430 and 5,960 respectively; of which 4,484 were in both databases and 2,946 were unique for PubMed and 1,476 were unique for Web of Science. Table 1 summarizes the major twenty types of diseases associated with epigenetics, and chemical compounds present in the food or natural products identified in the current search are listed. Table S1 in the Supporting Information summarizes the complete list associated with the respective related genes and epigenetic targets.

Table 1. Top twenty types of diseases associated with food epigenetic compounds.

Associated diseases	Epigenetic target
Breast cancer	DNMT1, DNMT3a, DNMT3b, HDAC1, HDAC2, HDAC3, HDAC 4, HDAC6, SIRT1, SIRT 2, SIRT 3, SIRT 4, SIRT5, SIRT6, KDM1B, KDM2A, KDM3A, KDM4A, KDM4B, KDM5A, KDM6B, KDM7A, KDM8
Lung cancer	DNMT1, DNMT3a, HDAC4, HDAC5, HDAC6, HDAC8, HDAC9, SIRT2, KDM1A, KDM3B
Prostate cancer	DNMT1, HDAC, HDAC4, HDAC5, HDAC6, KDM1A, KDM2B, SIRT1
Colorectal cancer	DNMT, HDAC7, KDM6B
Bladder cancer	HDAC6, LSD1, KDM6A
Melanoma	HDAC2, HDAC5, KDM5A, KDM6A
Oral cancer	HDAC6, HDAC8, KDM1A
Hepatocellular carcinoma	DNMT3a, HDAC10, KDM1A, KDM2A
Alzheimer's	DNMT, HDAC3, SIRT1
Endometrial cancer	DNMT, DNMT1, HDAC 3, KDM4A
Non-small cell lung cancer	DNMT3a, HDAC1, HDAC2, KDM6B
Gastric cancer	DNMT1, DNMT3a, HDAC 2, KDM2A, KDM2B
Cervical cancer	DNMT1, HAT/Ep300, HAT2B/Ep300, KDM5C
Colon cancer	DNMT3b, HDAC 1, HDAC 3, HDAC 7, KDM4C, KDM5A, KDM6B
Diabetes mellitus type 2	DNMT, HDAC, SIRT1
Glioblastoma	LSD1, KDM1A
Obesity and metabolic diseases	DNMT, HDAC1, SIRT1
Esophageal carcinoma	DNMT, HAT2B/Ep300
Squamous cell carcinoma	HDAC, HDAC5
Atherosclerosis	DNMT, HDAC7, SIRT1

3.2. Compound database

A total of 436 papers out of 8,906 unique papers from both databases (PubMed and Web of Science) were used as the basis to build and curate the compound data set introduced in this work. The current version of the data set contains 187 unique compounds, of which; 121 compounds have reported specific activity against at least one of the targets, and 66 compounds have reported general activity against at least one target family. The Epi Food Chemical Database contains ten columns with general information plus forty-nine columns that encode the epigenetic activity profile of the compounds across forty-six epigenetic targets. The general information is comprised of structural information in three linear notations, namely SMILES, InChi, and InChi keys; name, source of the compound, DOI of the peer-reviewed reference reporting the epigenetic activity, and links to FooDB and COCONUT databases through hyperlinks using the corresponding ID's on these two public databases.

The epigenetic activity profile is encoded as bit vectors of 0 and 1, indicating the absence or presence of reported activity, respectively, for each of the 46 targets (see the Methods section 2.2 for details). The epigenetic targets are ordered and arranged into three main groups: writers, erasers, and readers as follows: 8 writers (DNMT1, DNMT3a, DNMT3b, HAT/Ep300, HAT2B/Ep300, HAT3B/p300, EZH2, PRMT1); 37 erasers (HDAC1, HDAC2, HDAC3, HDAC4, HDAC5, HDAC6, HDAC7, HDAC8, HDAC9, HDAC10, HDAC11, SIRT1, SIRT2, SIRT3, SIRT4, SIRT5, SIRT6, SIRT7, LSD1, KDM1A, KDM1B, KDM2A, KDM2B, KDM3A, KDM3B, KDM4A, KDM4B, KDM4C, KDM4D, KDM5A, KDM5B, KDM5C, KDM5D, KDM6A, KDM6B, KDM7A, KDM8) and 1 reader (BET/BRD4). The main sources of the food chemicals in the Epi Food Chemical Database are meat, legumes, whole grains, grapes, poultry, acorn, acerola, strawberries, nuts, etc.

The 15 most frequent targets with reported activity of the compounds in the database are shown in Figure 3. We can see that the most frequent target is DNMT1 (63), followed by DNMT3B (35) and DNMT3A (34), HDAC6 (31) and HDAC1 (28).

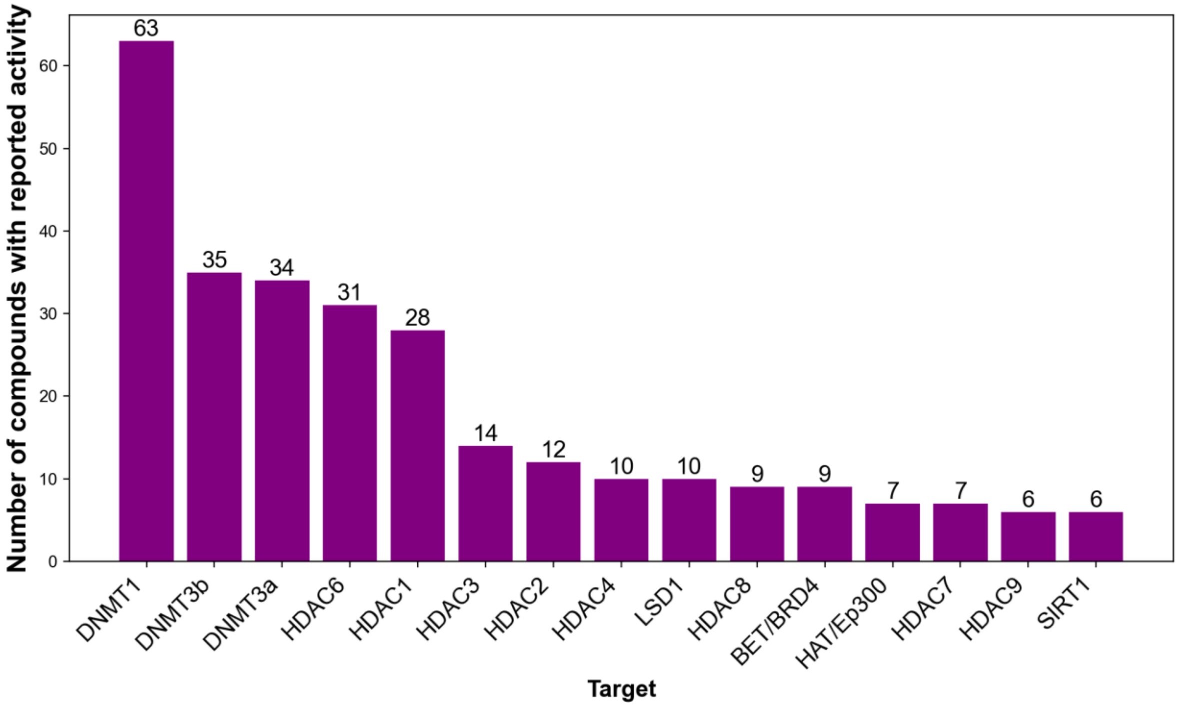

Figure 3. Histogram showing the 15 most frequent epigenetic targets.

There are 58 compounds with reported specific activity for only one target, being DNMT1 and HDAC6 the most frequent epigenetic targets with 18 compounds each one, followed by LSD1 with eight compounds, BET/BRD4 with four compounds and DNMT3a, DNMT3b, HAT/Ep300, KDM4a, with activity vs. two compounds in any case. Furthermor,e there are three epigenetic targets associated with specific reported activity vs only one compound each: HDAC1 with phenethyl isothiocyanate (PEITC), SIRT1 with pterostilbene and SIRT 5 with glutamate. The five compounds identified in the search with activity vs. the largest number of epigenetic targets were: biotin (27 targets), berberine (15 targets), alpha-ketoglutarate (13 targets), trichostatin (12 targets), and butein (11 targets). Additional compounds are shown in Figure 4, including the chemical structure and the number of targets in parenthesis.

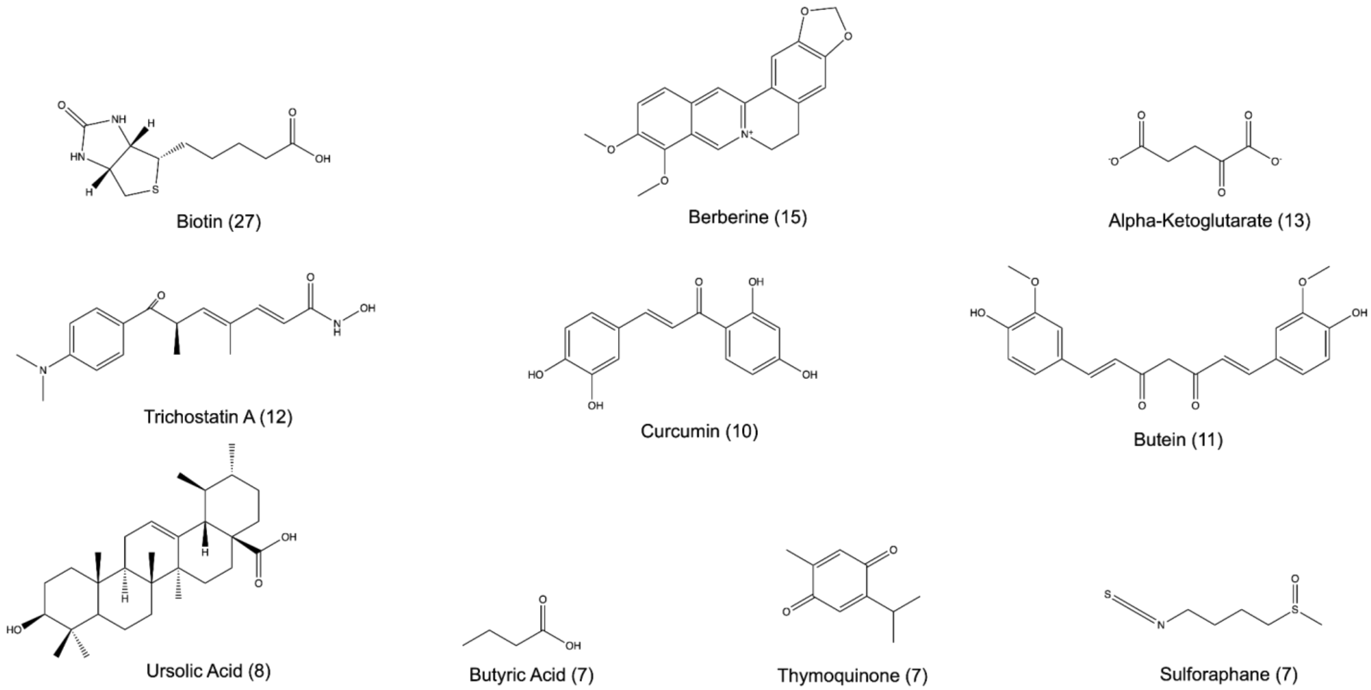

Figure 4. Top 10 chemical compounds in Epi Food Chemical Database with epigenetic activity.

3.3. Structural contents; diversity analysis and chemical and multiverse analysis

3.3.1. Diversity analysis

The total number of unique scaffolds for the 187 compounds was 91. Figure 5 shows the ten most frequent scaffolds along with the frequency and percent proportion, which represent 35.54% of the total distribution. The most frequent scaffolds were benzene (10.37%), followed by flavone (5.93%) and flavylium (2.96%). Other frequent scaffolds are indole (2.96%), pyridine (2.22%), hexane (2.22%), and isoflavone (1.48%).

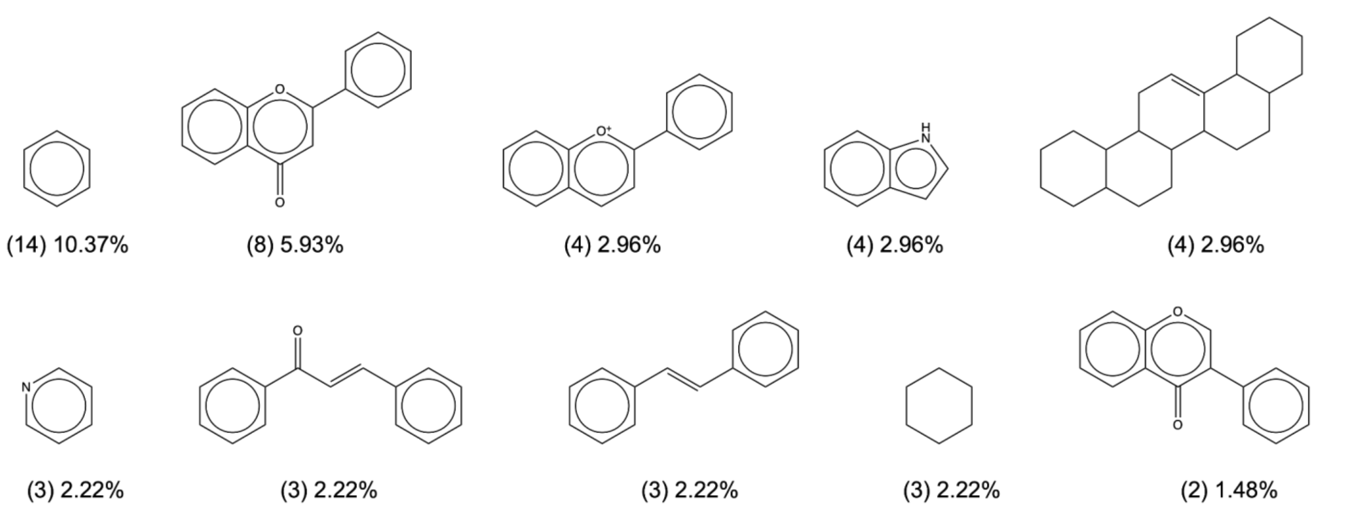

Figure 5. The ten most frequent scaffolds in Epi Food Chemical Database.

Figure 6 shows the cyclic system recovery (CSR) curve for the scaffold diversity in the Epi Food Chemical Database. This curve illustrates the proportion of molecules within a dataset that belong to a specific fraction of scaffolds. In a dataset with high diversity, each molecule in the library would correspond to a different scaffold, resulting in a diagonal with an AUC of 0.5. As the range of scaffold diversity diminishes, the curve will deviate from the diagonal orientation. Otherwise, the nadir of diversity would show in a dataset wherein all compounds share the same chemical scaffold; in such an instance, the CSR curve would appear as a vertical line, accompanied by an AUC of 1.0.[44] The shape of the CSR curve in Figure 6 indicates a large scaffold diversity.

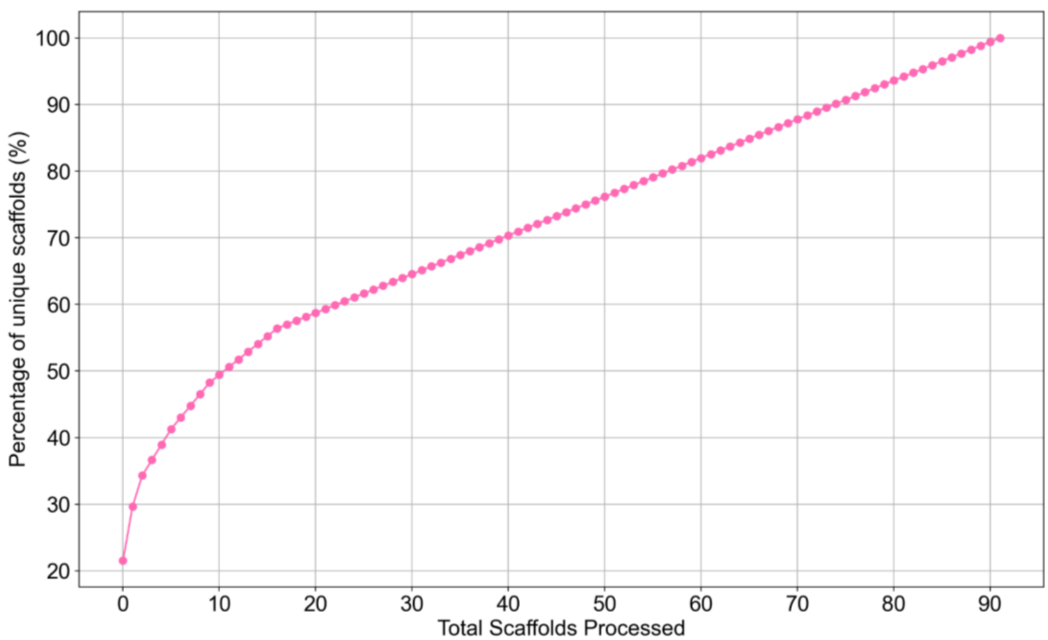

Figure 6. Cyclic system recovery curve of Bemis & Murcko scaffold diversity.

3.3.2. Visualization of chemical/molecular and multiverse spaces

The newly developed/constructed Epi Food Chemical Database with 187 chemical compounds in food/natural products was visualized in a graphical t-SNE representation of the chemical space. For comparison, FooDB was included as a reference in the visualization. The t-SNE was performed based on the physicochemical descriptors of the chemical compounds in the databases with the module *MoleculeDescriptors* of RDKit, such as: Molecular Weight, octanol/water coefficient (logP), number of hydrogen donor atoms (HBD), number of hydrogen acceptor atoms (HBA), topological polar surface area

(TPSA), number of aromatic heterocycles, number of aromatic rings, number of heteroatoms, number of rotatable bonds, etc.

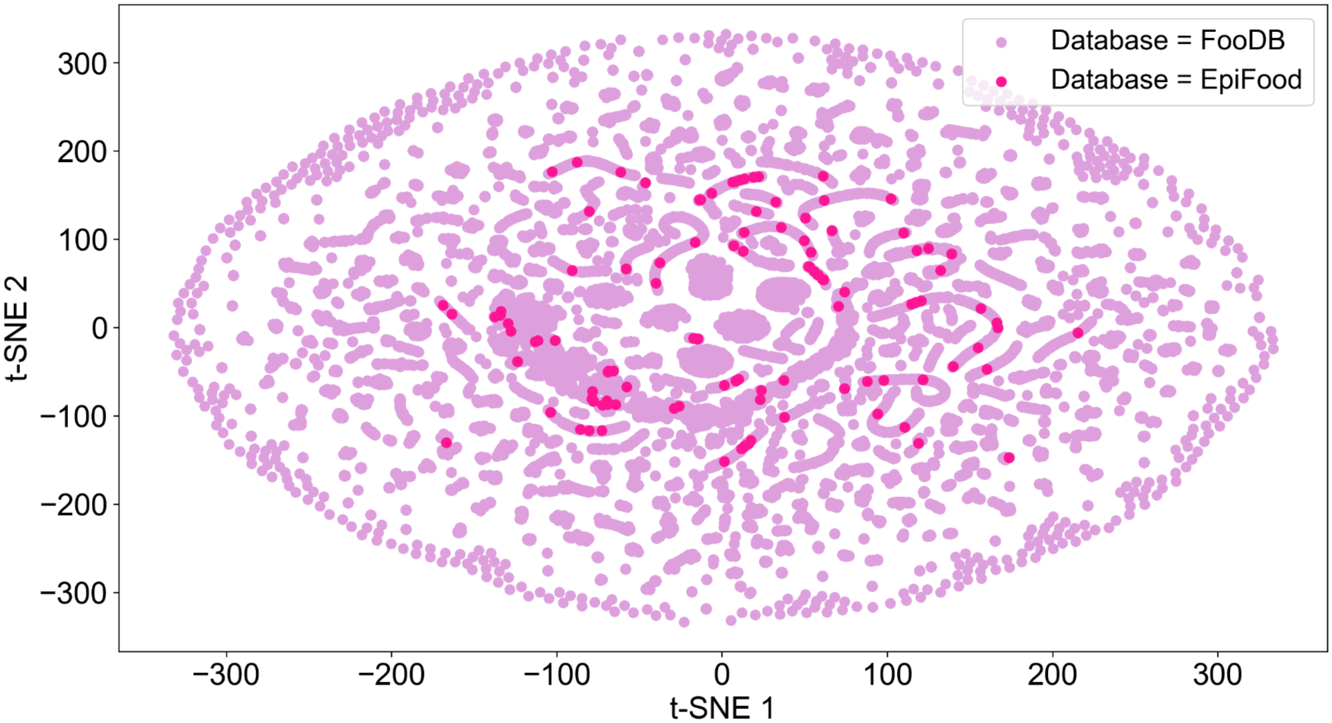

Figure 7. t-SNE showing the compounds in the database; in deep pink chemical compounds in Epi Food Chemical Database, in lilac chemical compounds in FooDB.

3.4. Structure-epigenetic target activity relationships

Figure 8 shows the SAS maps for the 187 chemical compounds in the Epi Food Chemical Database with the four different fingerprints: A) ECFP4, B) ECFP6, D) MACCS Keys, and D) RDKit fingerprint. The four interactive plots of the SAS maps are available in the Supporting information.

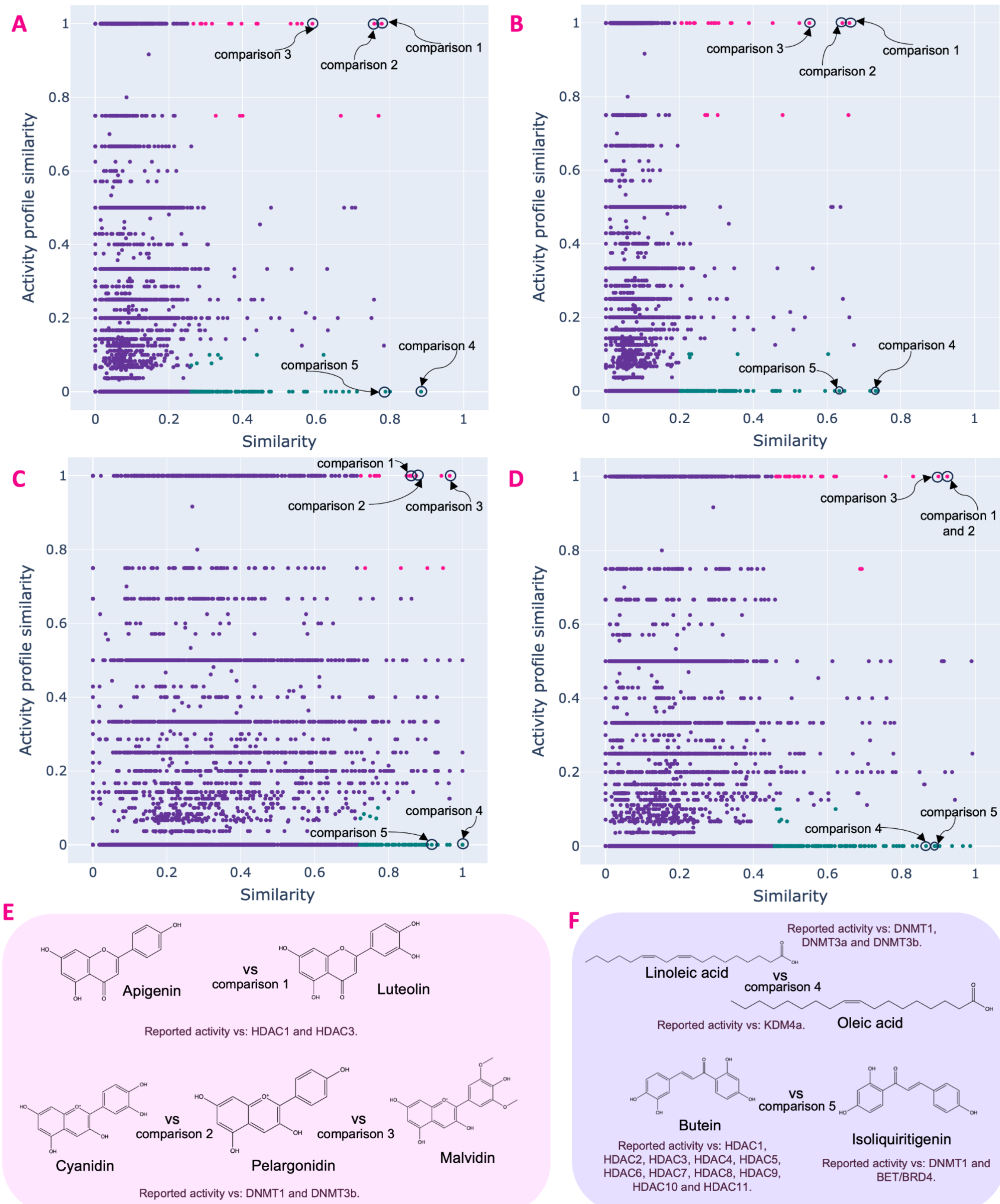

Figure 8. SAS maps, in pink are compounds in region II of the SAS map: similar structures and similar activity profile, in green are compounds in the IV region of the SAS map: similar activity profiles but very different activity profiles (activity cliffs). A) SAS map with ECFP4 fingerprint, B) SAS map with ECFP6 fingerprint, C) SAS map with MACCSKeys fingerprint, D) SAS map with RDKit fingerprint, E) examples of common compounds that are present in region II (pink points) of the four (A-D) SAS maps, F) examples of common compounds that are present in region IV (green points) of the four (A-D) SAS maps.

The pink data points represent the pair of chemical compounds present in region II of the SAS maps, which correspond to compounds very similar in structure as in profile activity. An example of this compound pair that is in common in the SAS maps of the four fingerprints is apigenin vs. luteolin (Figure 8)). These compounds have reported activity vs. HDAC1 and HDAC3, and some of the principal mains of both compounds are parsley, celery, onions, and pepper. Other examples of compounds in this region of the SAS maps are the comparisons between cyanidin vs. malvidin vs. pelargonidin; in this case, the compounds have reported activity vs. DNMT1 and DNMT3b, and some of the principal sources of the three compounds are blackberries, cherries, strawberries, and raspberries.

In contrast, the green data points represent a pair of compounds in region IV of the SAS maps, corresponding to compounds with similar activity profiles but very different structures. Examples of these pairs of compounds present in region IV of all SAS maps for all the fingerprints are linoleic acid with reported activity vs. DNMT1, DNMT3a, and DNMT3b and oleic acid with reported activity vs. KDM4, and their main sources are avocado, nuts, vegetable oils, and seeds. Another pair of compounds is butein with reported activity vs. HDAC1, HDAC2, HDAC3, HDAC4, HDAC5, HDAC6, HDAC7, HDAC8, HDAC9, HDAC10 and HDAC 11 vs. isoliquiritigenin with reported activity vs DNMT1 and BET/BRD4 in which their main sources are soybeans, peanuts, strawberries, and raspberries.

It is important to remember that the pairwise compounds of this work are based on the reported activity in the literature. For this reason, it is better to call them "pseudo activity cliffs" or pro-activity cliffs[45] instead of activity cliffs to the compounds in region IV. This is because maybe there are pairs of compounds that have very similar profile activity but have not been explored yet. Examples of these compounds are apigenin and luteolin vs. chrysin. With current data reported in the literature, it is concluded that apigenin and luteolin are compounds that have similar structures with the same activity profile with reported activity vs. HDAC1 and HDAC3, but both compounds are pseudo activity cliffs vs. chrysin, which have activity reported vs. HDAC6. So it is very probable that chrysin could have activity to HDAC1 and HDAC3 but also that apigenin and luteolin could also have activity vs. HDAC6.

4. Conclusions

Here we report the construction and curation of the Epi Food Chemical Database, which contains 187 chemical compounds from dietary and natural products. The database contains structural information and the epigenetic activity profile obtained from the literature vs. 46 epigenetic targets. We used chemoinformatic tools to compare and analyze the structural content, diversity, and chemical space. Scaffold analysis revealed that the most frequent scaffolds were benzene, followed by flavone and flavylium. In addition, we identified two main groups of compounds; the first, with continuous structure-activity relationships, aka, fulfill the similarity principle: compounds with similar chemical structures have similar epigenetic activity profiles. The second group of compounds can be considered pseudo-activity cliffs (similar structures but very different epigenetic activity profile). We suggest additional experimental testing of the compounds that form pseudo-activity cliffs. They may have similar activity to their corresponding compounds. This work contributes to the further advancement of a systematic analysis of food and natural product chemicals with epigenetic activity using chemoinformatic approaches.

Supporting Information:

The supporting information is available at GitHub https://github.com/EuridiceJuarez/EpiFoodChemicalDatabase. It contains the annotated compound database of food chemicals reported with epigenetic activity (Epi Food Chemical Database) in CSV format; Table S1 with the list of diseases/genes obtained in the literature search; Table S2 summarizing the list of 436 research papers used to build the Epi Food Chemical Database. and the interactive SAS maps plots of compounds in the Epi Food Chemical Database.

Acknowledgments

KE J-M, JF A-T, AL C-H thank *Consejo Nacional de Humanidades, Ciencias y Tecnologías* (CONAHCyT), Mexico, for the postgraduate scholarships 893849, 1270553, 847870. H V-Q thanks UNAM-HUAWEI for the scholarship under the project no. 7, “Desarrollo y aplicación de algoritmos de inteligencia artificial para el diseño de fármacos aplicables al tratamiento de diabetes mellitus y cáncer”.

Funding: DGAPA, UNAM, Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), grant No. IN201321.

References

1. Dupont C, Armant DR, Brenner CA. Epigenetics: definition, mechanisms and clinical perspective. Seminars in Reproductive Medicine. 2009;27(5):351-357. DOI: 10.1055/s-0029-1237423
2. Lacal I, Ventura R. Epigenetic Inheritance: Concepts, Mechanisms and Perspectives. Frontiers in Molecular Neuroscience. 2018;11. doi:10.3389/fnmol.2018.00292
3. Li Y. Modern epigenetics methods in biological research. Methods. 2021;187:104-113. DOI: 10.1016/j.ymeth.2020.06.022
4. al Theyab A, Almutairi T, Al-Suwaidi AM, Bendriss G, McVeigh C, Chaari A. Epigenetic Effects of Gut Metabolites: Exploring the Path of Dietary Prevention of Type 1 Diabetes. Frontiers in Nutrition. 2020;7. DOI: 10.3389/fnut.2020.563605
5. Ramírez-Alarcón K, Victoriano M, Mardones L, et al. Phytochemicals as Potential Epidrugs in Type 2 Diabetes Mellitus. Frontiers in Endocrinology. 2021;12. DOI: 10.3389/fendo.2021.656978
6. Carlos-Reyes Á, López-González JS, Meneses-Flores M, et al. Dietary Compounds as Epigenetic Modulating Agents in Cancer. Frontiers in Genetics. 2019;10. DOI: 10.3389/fgene.2019.00079
7. Aleksandrova K, Romero-Mosquera B, Hernandez V. Diet, Gut Microbiome and Epigenetics: Emerging Links with Inflammatory Bowel Diseases and Prospects for Management and Prevention. Nutrients. 2017;9(9):962. DOI: 10.3390/nu9090962
8. Milagro FI, Martinez JA. Dietary and Metabolic Compounds Affecting Covalent Histone Modifications. In: Handbook of Epigenetics. Elsevier; 2017:307-322. DOI: 10.1016/B978-0-12-805388-1.00020-1
9. Evans LW, Stratton MS, Ferguson BS. Dietary natural products as epigenetic modifiers in aging-associated inflammation and disease. Natural Product Reports. 2020;37(5):653-676. DOI: 10.1039/C9NP00057G
10. Miranda Furtado CL, dos Santos Luciano MC, Silva Santos R da, Furtado GP, Moraes MO, Pessoa C. Epidrugs: targeting epigenetic marks in cancer treatment. Epigenetics. 2019;14(12):1164-1176. DOI: 10.1080/15592294.2019.1640546
11. Jin, Y.; Liu, T.; Luo, H.; Liu, Y.; Liu, D. Targeting Epigenetic Regulatory Enzymes for Cancer Therapeutics: Novel Small-Molecule Epidrug Development. Frontiers in Oncology 2022, 12. DOI: 10.3389/fonc.2022.848221
12. Arce, C.; Segura-Pacheco, B.; Perez-Cardenas, E.; Taja-Chayeb, L.; Candelaria, M.; Dueñnas-Gonzalez, A. Hydralazine Target: From Blood Vessels to the Epigenome. Journal of Translational Medicine 2006, 4 (1). DOI: 10.1186/1479-5876-4-10

13. Yoon, S.; Eom, G. H. HDAC and HDAC Inhibitor: From Cancer to Cardiovascular Diseases. Chonnam Medical Journal 2016, 52 (1), 1. DOI: 10.4068/cmj.2016.52.1.1

14. Bennett, R. L.; Licht, J. D. Targeting Epigenetics in Cancer. Annual Review of Pharmacology and Toxicology 2018, 58 (1), 187–207. DOI: 10.1146/annurev-pharmtox-010716-105106

15. Meroni M, Longo M, Rustichelli A, Dongiovanni P. Nutrition and Genetics in NAFLD: The Perfect Binomium. International Journal of Molecular Sciences. 2020;21(8):2986. DOI: 10.3390/ijms21082986

16. González-Becerra K, Ramos-Lopez O, Barrón-Cabrera E, et al. Fatty acids, epigenetic mechanisms and chronic diseases: a systematic review. Lipids in Health and Disease. 2019;18(1):178. DOI: 10.1186/s12944-019-1120-6

17. Ali A, Hamzaid NH, Ismail NAS. The Interplay of Nutriepigenomics, Personalized Nutrition and Clinical Practice in Managing Food Allergy. Life. 2021;11(11):1275. DOI: 10.3390/life11111275

18. Li Y, Buckhaults P, Li S, Tollefsbol T. Temporal efficacy of a sulforaphane-based broccoli sprout diet in prevention of breast cancer through modulation of epigenetic mechanisms. Cancer Prevention Research. 2018;11(8):451-464. DOI: 10.1158/1940-6207

19. Regal P, Fente C, Cepeda A, Silva E. Food and omics: unraveling the role of food in breast cancer development. Current Opinion in Food Science. 2021;39:197-207. DOI: 10.1016/j.cofs.2021.03.008

20. Cuyàs E, Castillo D, Llorach-Parés L, et al. Computational de-orphanization of the olive oil biophenol oleacein: Discovery of new metabolic and epigenetic targets. Food and Chemical Toxicology. 2019;131:110529. DOI: 10.1016/j.fct.2019.05.037

21. Taniguchi T, Tischkowitz M, Ameziane N, et al. Disruption of the Fanconi anemia–BRCA pathway in cisplatin-sensitive ovarian tumors. Nature Medicine 2003 9:5. 2003;9(5):568-574. DOI: 10.1038/nm852

22. Medina-Franco, J. L.; Saldívar-González, F. I. Cheminformatics to Characterize Pharmacologically Active Natural Products. Biomolecules 2020, 10 (11), 1566. DOI: 10.3390/biom10111566.

23. Martinez-Mayorga, K.; Medina-Franco, J. L. Foodinformatics. Applications of Chemical Information to Food Chemistry 2014. DOI: 10.1007/978-3-319-10226-9

24. López-López, E.; Bajorath, J.; Medina-Franco, J. L. Informatics for Chemistry, Biology, and Biomedical Sciences. Journal of Chemical Information and Modeling 2020, 61 (1), 26–35. DOI: 10.1021/acs.jcim.0c01301

25. Kirchmair, J. Molecular Informatics in Natural Products Research. Molecular Informatics 2020, 39 (11), 2000206. DOI: 10.1002/minf.202000206

26. Medina-Franco, J. L.; Saldívar-González, F. I. Cheminformatics to Characterize Pharmacologically Active Natural Products. Biomolecules 2020, 10 (11), 1566. DOI: 10.3390/biom10111566

27. http://www.foodb.ca/

28. Sorokina, M.; Merseburger, P.; Rajan, K.; Yirik, M. A.; Steinbeck, C. Coconut Online: Collection of Open Natural Products Database. Journal of Cheminformatics 2021, 13 (1). DOI: 10.1186/s13321-020-00478-9

29. Medina-Franco, J. L.; Chávez-Hernández, A. L.; López-López, E.; Saldívar-González, F. I. Chemical Multiverse: An Expanded View of Chemical Space. Molecular Informatics 2022, 41 (11). DOI: 10.1002/minf.202200116

30. Maggiora, G.; Medina-Franco, J. L.; Iqbal, J.; Vogt, M.; Bajorath, J. From Qualitative to Quantitative Analysis of Activity and Property Landscapes. Journal of Chemical Information and Modeling 2020, 60 (12), 5873–5880. DOI: 10.1021/acs.jcim.0c01249

31. PubMed. https://pubmed.ncbi.nlm.nih.gov/

32. Web of Science Core Collection. https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/web-of-science-core-collection/

33. Weininger, D. Smiles, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. Journal of Chemical Information and Computer Sciences 1988, 28 (1), 31–36. DOI: 10.1021/ci00057a005

34. Maggiora, G.; Medina-Franco, J. L.; Iqbal, J.; Vogt, M.; Bajorath, J. From Qualitative to Quantitative Analysis of Activity and Property Landscapes. Journal of Chemical Information and Modeling 2020, 60 (12), 5873–5880. DOI: 10.1021/acs.jcim.0c01249

35. Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. Journal of Medicinal Chemistry 1996, 39 (15), 2887–2893. DOI: 10.1021/jm9602928

36. Chung, N. C.; Miasojedow, B.; Startek, M.; Gambin, A. Jaccard/Tanimoto Similarity Test and Estimation Methods for Biological Presence-Absence Data. BMC Bioinformatics 2019, 20 (S15). DOI: 10.1186/s12859-019-3118-5

37. Kang, B.; García García, D.; Lijffijt, J.; Santos-Rodríguez, R.; De Bie, T. Conditional T-SNE: More Informative T-SNE Embeddings. Machine Learning 2020, 110 (10), 2905–2940. DOI: 10.1007/s10994-020-05917-0

38. Zhou, Y.; Sharpee, T. O. Using Global T-SNE to Preserve Intercluster Data Structure. Neural Computation 2022, 34 (8), 1637–1651. DOI: 10.1162/neco_a_01504

39. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. Journal of Chemical Information and Modeling 2010, 50 (5), 742–754. DOI: 10.1021/ci100050t

40. Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. Characterization of Activity Landscapes Using 2D and 3D Similarity Methods: Consensus Activity Cliffs. Journal of Chemical Information and Modeling 2009, 49 (2), 477–491. DOI: 10.1021/ci800379q

41. Medina-Franco, J. L. Scanning Structure–Activity Relationships with Structure–Activity Similarity and Related Maps: From Consensus Activity Cliffs to Selectivity Switches. Journal of Chemical Information and Modeling 2012, 52 (10), 2485–2493. DOI: 10.1021/ci300362x

42. Waddell, J.; Medina-Franco, J. L. Bioactivity Landscape Modeling: Chemoinformatic Characterization of Structure–Activity Relationships of Compounds Tested across Multiple Targets. Bioorganic & Medicinal Chemistry 2012, 20 (18), 5443–5452. DOI: 10.1016/j.bmc.2011.11.051

43. Medina-Franco, J. L.; Navarrete-Vázquez, G.; Méndez-Lucio, O. Activity and Property Landscape Modeling Is at the Interface of Chemoinformatics and Medicinal Chemistry. Future Medicinal Chemistry 2015, 7 (9), 1197–1211. DOI: 10.4155/fmc.15.51

44. Lipkus, A. H.; Watkins, S. P.; Gengras, K.; McBride, M. J.; Wills, T. J. Recent Changes in the Scaffold Diversity of Organic Chemistry as Seen in the CAS Registry. Journal of Organic Chemistry 2019, 84 (21), 13948–13956. DOI: 10.1021/acs.joc.9b02111

45. Waddell, J.; Medina-Franco, J. L. Bioactivity Landscape Modeling: Chemoinformatic Characterization of Structure–Activity Relationships of Compounds Tested across Multiple Targets. Bioorganic & Medicinal Chemistry 2012, 20 (18), 5443–5452. DOI: 10.1016/j.bmc.2011.11.051