Published in final edited form as:

Mol Pharm. 2018 March 05; 15(3): 840–850. doi:10.1021/acs.molpharmaceut.7b00412.

A Comprehensive Database and Analysis Framework To Incorporate Multiscale Data Types and Enable Integrated Analysis of Bioactive Polyphenols

Lap Ho^{†,‡,§}, Haoxiang Cheng^{||,⊥,§}, Jun Wang[†], James E. Simon[#], Qingli Wu[#], Danyue Zhao[#], Eileen Carry[¶], Mario G. Ferruzzi[§], Jeremiah Faith^{||}, Breanna Valcarcel[†], Ke Hao^{||,⊥}, and Giulio M. Pasinetti^{†,‡,*}

[†]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States

^{||}Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States

[⊥]Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States

[‡]Geriatric Research, Education & Clinical Center, James J. Peters Veterans Affairs Medical Center, Bronx, New York 10468, United States

[#]New Use Agriculture and Natural Plant Products Program, Department of Plant Biology, Rutgers University, New Brunswick, New Jersey 08901, United States

[¶]Department of Medicinal Chemistry, Ernest Mario School of Pharmacy, Piscataway, New Jersey 08854, United States

[§]Plants for Human Health Institute, North Carolina State University, Kannapolis, North Carolina 27695, United States

Abstract

The development of a given botanical preparation for eventual clinical application requires extensive, detailed characterizations of the chemical composition, as well as the biological availability, biological activity, and safety profiles of the botanical. These issues are typically addressed using diverse experimental protocols and model systems. Based on this consideration, in this study we established a comprehensive database and analysis framework for the collection, collation, and integrative analysis of diverse, multiscale data sets. Using this framework, we

*Corresponding Author: giulio.pasinetti@mssm.edu.

ORCID

Mario G. Ferruzzi: 0000-0003-1384-2672

Giulio M. Pasinetti: 0000-0002-1524-5196

§Author Contributions

L.H. and H.C. contributed equally to this work.

Notes

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.molpharmaceut.7b00412. (PDF)

conducted an integrative analysis of heterogeneous data from *in vivo* and *in vitro* investigation of a complex bioactive dietary polyphenol-rich preparation (BDPP) and built an integrated network linking data sets generated from this multitude of diverse experimental paradigms. We established a comprehensive database and analysis framework as well as a systematic and logical means to catalogue and collate the diverse array of information gathered, which is securely stored and added to in a standardized manner to enable fast query. We demonstrated the utility of the database in (1) a statistical ranking scheme to prioritize response to treatments and (2) in depth reconstruction of functionality studies. By examination of these data sets, the system allows analytical querying of heterogeneous data and the access of information related to interactions, mechanism of actions, functions, etc., which ultimately provide a global overview of complex biological responses. Collectively, we present an integrative analysis framework that leads to novel insights on the biological activities of a complex botanical such as BDPP that is based on data-driven characterizations of interactions between BDPP-derived phenolic metabolites and their mechanisms of action, as well as synergism and/or potential cancellation of biological functions. Our integrative analytical approach provides novel means for a systematic integrative analysis of heterogeneous data types in the development of complex botanicals such as polyphenols for eventual clinical and translational applications.

Graphical Abstract

Keywords

multiple scale data set; integrated analysis; directional networks; dietary botanicals and polyphenol metabolites

INTRODUCTION

Epidemiological studies and associated meta-analyses strongly suggest that long-term consumption of diets rich in plant polyphenols may promote healthy aging¹ and protect against a variety of diseases, including cancer,^{2,3} cardiovascular disease,^{4,5} and neurodegenerative disorders.^{6–8} Polyphenols belong to a structural class of organic compounds characterized by the presence of two or more phenol structural units, which are aromatic organic structures with the base molecular formula C_6H_5OH . Polyphenols are found abundantly in fruits such as grapes, berries, vegetables, tea, and a wide array of plants.⁹ The health benefits of dietary polyphenols have been commonly attributed to their direct and indirect antioxidant and anti-inflammatory characteristics.^{10,11} However, more recent evidence suggests that dietary polyphenols may also benefit those who suffer from diverse diseases by protecting against disease-specific pathogenic processes.^{12–16}

In spite of supportive evidence, the development of specific polyphenol-rich dietary or supplemental preparations from specific plant sources for disease prevention and/or treatment is complicated by significant differences in the chemical composition of these preparations generated from a given plant source.^{17,18} The exact chemical composition of individual botanical-based natural product mixtures is influenced by many factors, including the genetics of the plant as well as differences in the locale and conditions of plant cultivation, plant harvesting, and processing into specific dietary preparations, storage/distributions of these dietary preparations, and the manners in which they are applied, which can all impact the chemical composition.^{19,20} Moreover, polyphenols in dietary preparations are typically not physiologically available *in vivo* in their native forms. Instead, the majority of orally consumed dietary polyphenols are extensively metabolized in the gastrointestinal (GI) track by GI bacteria, during GI absorption, and/or by postabsorptive xenobiotic metabolism, converting them into metabolite forms that are typically biologically available *in vivo*, either through further substitution of the polyphenol nucleus (e.g., via methylation, glucuronidation, or sulfation) or by degradation (e.g., to phenolic acids).^{21–24} Thus, putative health benefits of bioactive polyphenol-rich dietary preparations are primarily mediated by biologically available phenolic metabolites that are derived from these dietary preparations. Successful preclinical and clinical development of a given bioactive polyphenol-rich dietary preparation for eventual clinical and translational applications therefore requires a comprehensive understanding of not only the chemical composition characteristics of the dietary preparation under development, but also the chemical consistency and reproducibility of the dietary preparation, the dose-responsive biological activity and safety profile of the dietary preparation, the pharmacokinetics and biological availability of phenolic metabolites from the dietary preparation, and the biological activities as well as interactions among these biologically available phenolic metabolites in specific target tissue(s) that may contribute to mechanistic modulation of disease-specific pathogenic processes. Most of these issues are typically investigated using different *in vitro* and *in vivo* model systems, different test reagents (e.g., a dietary polyphenol preparation, or select polyphenol compound or phenolic metabolite from the polyphenol preparation), diverse test doses, and varying experimental protocols. Systematic integrations of these heterogeneous data sources and types are critical for the preclinical and clinical development of a given dietary polyphenol preparation, yet have been challenging to accomplish.

We recently identified and extensively characterized a bio-active dietary polyphenol-rich preparation (BDPP). We found BDPP to be highly effective in protecting against the onset and/or progression of multiple, diverse neurological, psychological, and metabolic disorders in animal models.^{25–27} BDPP is composed of a select Concord grape juice (CGJ), a select grape seed polyphenol extract (GSPE), and *trans*-resveratrol (RSV),^{25–27} which provides a complex mixture of diverse naturally occurring polyphenols.^{17,18,25} We have gathered extensive published^{25–27} and unpublished information on the effects of oral BDPP administration in modulating multiple pathologic behavioral phenotypes and/or cellular/molecular pathways across diverse animal models of cognitive, psychological, and metabolic disorders. We have also gathered extensive published^{13,25,26,28–30} and unpublished information regarding the pharmacokinetics and bioavailability of BDPP or individual BDPP polyphenol components, including their distribution to peripheral tissues and specifically to

the central nervous system. Moreover, we have gathered extensive published^{13,25–27,29,31–34} and unpublished observations on the biological activity, drug-like properties, and safety profiles of many of the biologically available BDPP-derived phenolic metabolites using a multitude of diverse *in vitro* and/or *in vivo* experimental models. In view of the heterogeneous BDPP data types we have gathered from different sources, and the need to link diverse aspects of these findings for translation, the overall goal of this study is to establish a comprehensive database and analysis framework as a systematic and logical means to catalogue and collate the diverse information we have gathered to date, and will continue to gather, to derive a modeling system that will allow us to provide a unified view of BDPP's bioavailability and bioactivity. This analysis framework will ultimately allow for the assessment of how BDPP and its bioactive metabolites may connect and interact with each other to mechanistically explain modulation of select disease-specific pathogenic pathways and impact the health benefits of BDPP. We hypothesize that such an analytical framework can easily be adapted for integrative analysis toward translational development of other complex bioactive botanical preparations.

EXPERIMENTAL SECTION

Data Types and Sources

As we have described in the Introduction, the data sets we utilized for the development of our comprehensive database and analysis framework include diverse published and unpublished data we have gathered from BDPP, complex dietary components (e.g., CGJ, GSPE) of BDPP, extracts (e.g., GSPE extracts containing polyphenols that are low, moderate, or high in molecular weight) and isolated polyphenol components (e.g., catechin, epicatechin, quercetin, etc.) from BDPP as well as its primary dietary components, and BDPP-derived phenolic biological metabolites, including polyphenolic metabolites (e.g., conjugated metabolites such as 3-OMe-epicatechin-5-I-glucuronide, and quercetin-glucuronide) and phenolic acids (e.g., 3-hydroxybenzoic acid and 3-(3'-hydroxyphenyl)propionic acid). These data sets include information we have obtained from diverse *in vivo* studies using different experimental animal models (e.g., transgenic A β mouse model of Alzheimer's disease, social defeat-mediated mouse model of depression/anxiety, sleep deprivation-mediated mouse model of cognitive dysfunction, etc.) and *in vitro* studies using a multitude of heterogeneous experimental model systems involving primary as well as stable cell cultures (e.g., primary neuron cultures, monocyte cultures, etc.), organotypic cultures (e.g., brain slice cultures), tissue and cell extracts (e.g., mitochondria extract), and biochemical substrates (e.g., monomeric beta-amyloid peptides). In *in vivo* studies, we also used different experimental animal models (e.g., transgenic A β mouse model of Alzheimer's disease, social defeat-mediated mouse model of depression/anxiety, sleep deprivation-mediated mouse model of cognitive dysfunction, etc.). Each of our *in vivo* studies involved treating select experimental models with a specific dose (or doses) of BDPP, BDPP components, polyphenols, or phenolic metabolites that were applied under varying administration routes (e.g., through diet or drinking water, via intraperitoneal or intravenous injections or via oral gavage) and treatment protocols (e.g., acute vs chronic administration and acute on chronic administration). A variety of outcome measures were monitored, dependent on experimental design. For example, in some of our *in vivo* studies

we monitored for treatment effects on behavioral indices of cognitive/psychological functions and/or the onset/progress of select disease pathologic phenotypes. We also conducted *in vivo* pharmacokinetics studies in mouse and rat models to identify and characterize specific BDPP-derived phenolic metabolites that are biologically available and are capable of accumulation in peripheral tissues and in the brain.

Moreover, we also monitored for treatment safety/tolerability of BDPP doses by assessing animals' food/water intake, body weight, general behavioral profile, blood enzymes, and pathologic evaluations, among others. For our *in vitro* studies, we treated our experimental models with phenolic metabolites derived from BDPP (or BDPP components) that are biologically available in peripheral and/or brain tissues. The testing protocols, with respect to specific phenolic metabolites being tested, the test doses, and the timing of treatment all varied across studies. The outcomes of *in vitro* measurements included cellular biological responses to treatments (e.g., cellular changes in gene expression at mRNA or protein levels), organotypic responses to treatment (e.g., changes in electrophysiological responses and gene expression in brain slices), indices of phenolic metabolite toxicity, drug-like properties and stability (e.g., cytochrome C inhibition and binding in plasma or the brain, as well as half-life assessments), and biochemical modulations of select disease-relevant molecules (e.g., interference with aggregation of beta-amyloid or tau into neurotoxic soluble high-molecular-weight aggregate forms). Collectively, our integrative analysis was conducted using multiscale data types with a large degree of heterogeneity that are generated from varying sources. Moreover, these data are available in a variety of formats, layouts, levels of granularity, and vocabulary.

Overall Approach

The development of a comprehensive database and analysis framework was implemented in three phases, which we completed in order: (1) Data Collection and Collation, (2) Sample Level Data Analysis, and (3) Summarization, Integrative Data Analysis, and Interpretation (Figure 1).

Phase 1: Data Collection and Collation—The first phase of the study involved data collection and collation, which was subdivided into three steps: Database schema definition, standardization of nomenclature, and quality control. These steps are described in more detail below.

Relational Database Schema Definition: The first step in data collection required us to define a schema which could accommodate data sets of such variety reasonably well. Defining the schema was an iterative process which required close collaboration and interaction between personnel from a bio-informatics facility with personnel from wet laboratories who generated the data. After several iterations, and as we covered a sufficiently large subset of the full data set, we were able to establish a user-friendly database management system software for data collection, storage, and retrieval. To facilitate prototyping of data entry and schema, as well as sharing and bulk-editing by all users, the overall database schema is designed and constructed as a comprehensive database containing seven tables that allow for the collection and storage of the following information from each

and every study: (1) experiment, (2) sample (experimental model systems), (3) treatment information, (4) readout measurements defined by specific nomenclatures, (5) readout dictionary describing the meaning of these readout nomenclatures, (6) known relationships among the experimental/treatment paradigms, readouts, and their biological significances, and (7) target coefficients. Each of the seven tables is linked through consistent identifiers.

The most fundamental table is the “readout measurements” table, which collects all the experimental readouts collected across all the experiments. This table has the following required variables: experiment identifier, sample identifier, readout identifier, tissue, time, replicate, and value. Variables in “readout measurements” tables are further annotated in other tables for detailed info. For example, the experiment and sample identifiers each link to separate tables (“experiment” and “sample”) providing additional details about the experiment and samples, respectively. Following data entry, consistency of the data value and format was automatically checked, and such consistency check is performed after each data revision/update. Inconsistencies were flagged and reported for human review and correction. This overall schema is designed to allow for minor adaptations and evolutions as new data are available for consolidation into the overall database.

Standardization of Nomenclature across Experiments: Since the multitude of experiments in the database were performed by different teams/personnel at different times and in different locations, diverse nomenclatures are often used to annotate/describe the same experimental system, test compounds, readout parameters, and other information. Thus, during the construction of our analytical framework and collection of information into this framework, personnel from our bioinformatics team and the research laboratories met regularly to standardize all nomenclature usage during and after the data entry process.

Data Quality Control: The overall database undergoes a quality control review after each new entry. First, we examine if the data table follows predefined rules, and that nomenclatures are standardized. Second, we examine if the values are in possible range (in terms of biology and instrument). Third, we developed a reporting tool that produces several summary statistics of the data to help highlight potential issues (e.g., outliers). Summary statistics included a table recapitulating the main features of each experiment, such as number of samples, tissues, time points, and readouts, as well as the distribution of the numerical and qualitative measurements, through histograms, bar plots, and summary tables, which are further stratified by tissue and time points whenever appropriate. Data entries showing errors in any of these three steps are flagged and reported to the researcher for manual review so that amendments could be created when necessary. The database might go through several cycles of report generation followed by report reviews before passing the QC.

Phase 2: Statistical Analysis of Sample Level Data—For each of the studies in the overall database, associations between treatment factors and readouts were quantified using a robust linear regression formula: $E[y] = \mathbf{X}\beta$, where $\mathbf{X}$ is the design matrix of dimension $n \times p$, with n being the sample size and p being the number of parameters in the vector β (vector of coefficient). This analysis was executed using the robust linear regression implemented in the “rlm” function of the MASS R package, Version 7.3-45 (CRAN-

Package MASS free statistical software; <https://cran.r-project.org/web/packages/MASS/MASS.pdf>), with the “M” method and the default “Huber” psi function. The design matrix $\mathbf{X}$ was built according to the experimental design. Binary readouts were recoded as 0/1 variables. Confidence intervals on β were derived through a bootstrap procedure, while significance p -values were obtained through permutation testing. We used $b = 1,000$ replicates in both procedures. For defining the design matrix $\mathbf{X}$ and linking modeling results to quantities of interest, we identified 5 experimental setups: (1) treated vs control, (2) drug screening, (3) dose–response, (4) factorial, additive, and (5) factorial, full. Each experimental design is described below.

Treated vs Control Experiments: These experiments contain a control (nontreated) group and a treated group (or multiple treated groups) in which select test treatment modalities (e.g., test compounds) were administered as independent variables. In the case where there is one control and one treated group, the design matrix $\mathbf{X}$ has 2 columns, which we coded for an intercept and an indicator variable representing with or without treatment, respectively. The coefficient of the indicator variable (β_1) is interpreted as the average variation in the readout, y , in the presence of the tested treatment. Significance of the coefficient β_1 represents the significance of the tested treatment affecting the readout y .

Drug Screening Experiments: Our overall database also contains inputs from studies having one control and two or more treated groups with each of the treatment groups receiving a different treatment paradigm (e.g., drug screening experiments with multiple treatment groups and each treatment group receiving a different preparation of the same drug or different drug formulation). In this case, the design matrix $\mathbf{X}$ has a first intercept column and two or more treatment columns, each serving as an indicator of a treatment variable K , and the indicator variable is termed I_k . We calculate the coefficient associated with each indicator I_k , which represents the readout’s variation attributable to treatment K .

Dose Response Experiments: In dose response experiments, there is one control group and there are 2 or more treated groups to which increasing doses of the same compound were administered. We coded the design matrix $\mathbf{X}$ as an intercept term plus k indicator variables, one for each non-zero test dose. The coefficient β_k is interpreted as the average variation of the readout at dose k , compared to the control group. This formulation does not impose a functional form on the relationship between the treatment compound doses and the readout. Overall significance of the association between treatment and readout was quantified as the minimum p -value across the Wald tests of the coefficients β_k , $k = 1, \dots, K$.

Additive Factorial Experiments: In additive factorial experiments, 2 treatments A and B are tested in three different configurations across three experimental groups: a control group, to which neither treatment is administered; a treatment group TA (or TB), to which only treatment A (or treatment B) is administered; and a treatment group TAB, to which both treatments A and B are administered simultaneously. The design matrix $\mathbf{X}$ has 3 columns: an intercept, an indicator of treatment A, and an indicator of treatment B. Corresponding coefficients β_A and β_B are interpreted as variations in the readout y in the presence of either treatment. In this model, the combined effect of the 2 treatments is assumed to be additive.

Significance for each coefficient represents the significance of each treatment affecting the readout y .

Full Factorial Experiments: In full factorial experiments, 2 treatments A and B are tested in 4 different configurations across 4 experimental groups: a control group that received no treatment, a treatment group TA that received treatment A, a treatment group TB that received treatment B, and a treatment group TAB that received both treatments A and B simultaneously. The design matrix $\mathbf{X}$ has 4 columns: an intercept, an indicator I_A of treatment A, an indicator I_B of treatment B, and the product of the two indicators $I_{AB} = I_A \times I_B$. Coefficients β_A and β_B are interpreted as the variation in the readout y in the presence of treatments A and B, respectively. β_{AB} represents the synergistic effect between the 2 treatments, and is the additional average variation in the readout y on top of the additive effects of the 2 treatments, when the 2 treatments A and B are administered simultaneously. The significance of β_A and β_B represents, respectively, the significance of the effects of treatments A and B on readout y . The significance of β_{AB} represents the significance of the synergistic effect of the 2 treatments on the readout.

As sample level data, whenever applicable, were collected for different tissues, at different time points, and from different readouts, separate and independent models were estimated for each combination of tissue, time point, and readout. The end result of the statistical analysis was thus the collection of records: identifications and descriptions of the experiments, description of the experimental models and samples, description of treatment paradigms, readout information/observations, and association between treatments and readouts. These records are stored in a flat table format. Using this table, researchers can directly rank treatments based on results of multiple experiments to gather mechanistic insights on the biological activities of orally administered polyphenol-rich preparations and prioritize new investigations for further study.

Phase 3: Integrative Data Analysis and Interpretation—As illustrated in Figure 1, the overall database contents were distilled into two main categories: (1) Extensive collection of experimental data from numerous experimental types/sources that were typically derived to address impacts of a specific treatment paradigm (e.g., BDPP administration) on select outcomes (e.g., bioavailability and metabolism of select polyphenolic compounds, behavior/biochemical effects, safety/toxicity evidence, etc.). (2) Statistical data analysis of sample level measurements (e.g., effect of BDPP treatment on IL-6 plasma levels, effects of malvidin-glucoside treatment on IL-6 plasma levels in cultured blood cells, etc.).

Sample level information relating specific treatment paradigms and outcome measures (which we refer to as expert knowledge) was directly stored in a table of “expert knowledge”. The information was then organized in the form of ordered triplets: (molecule 1, relationship, molecule 2), where the relationship is the expert knowledge. Moreover, statistical analysis results were also stored in a triplet, (molecule 1, relationship, molecule 2), where the relationship is the significant statistical association (p -value ≤ 0.01) between the level/dose of molecules 1 and 2, and the relationship is also annotated for positive or

negative correlation, indicating whether the treatment increases or decreases the readout value, respectively

The collection of triplets (from expert knowledge, individual experiments, or summations from multiple experiments) we generated from the overall data set served as “building blocks”, which we then organized into a network. The construction was restricted by the following rules: (1) for each triplet, molecules were presented as network nodes, and the relationship was presented as a network connection; (2) each node can appear only once in the network; (3) the directionality of the network connection was presented as arrows, based on study design or prior knowledge; (4) the node layout is a hierarchical order connecting nodes from the top to the bottom of the graph (i.e., the directionality of most edges was downward); (5) node/connect overlap is minimized to improve readability. This network construction was carried out using a free statistical software, *igraph*, Version 1.0.1 (CRAN, Web site: <https://cran.r-project.org/web/packages/igraph/index.html>). Annotation was added to the network graph using *Cytoscape* software (Version 3.4.0, <http://cytoscape.org/>). The standard error (SE) was not directly used in the integrative analysis. Despite this, the *p*-value of the association was a function of the SE, and *p*-values were used in building the network connections. The network is constructed such that links (i.e., connections) in the networks present prior knowledge or statistically significant associations. The directionality of the connections (presented as an arrow) stands for BDPP components (prior knowledge), metabolites of BDPP *in vivo* (prior knowledge), or statistical associations (data driven). And the causal direction of the statistical association is determined by study design. When the network shows no connection between two nodes, it could be (1) no prior knowledge and no experiments done to examine the relationship of the nodes, or (2) no prior knowledge and experiments done to examine the statistical significance of the relationship.

RESULTS

We have recently identified a select bioactive dietary polyphenol preparation, referred to as BDPP, that, when administered as a dietary supplement, is effective in protecting against Alzheimer’s disease^{26,30,32–34} and metabolic syndrome²⁵ phenotypes in experimental models. More recently, we also demonstrated that BDPP is effective in promoting resilience against environmental stress-mediated cognitive dysfunction (e.g., sleep-deprivation-induced cognitive dysfunction²⁷) or psychological dysfunction (e.g., repeated social defeat stress-mediated depression/anxiety phenotypes⁵⁰). A better understanding of the mechanism by which BDPP may promote distinct health benefits is critical for the eventual clinical translation of BDPP.

Development of an Integrative Analytical Framework

We developed a comprehensive database and analysis framework to facilitate systematic collection and statistical integration of diverse, heterogeneous data types we have generated from BDPP experiments. The development of this framework is divided into three main phases, which we completed in order: (1) Data Collection and Collation, (2) Sample Level Data Analysis, and (3) Summarization, Integrative Data Analysis, and Interpretation. These three stages are schematically presented in Figure 1.

Data Collection and Collation

For Data Collection and Collation, we established a user-friendly database for data collection, storage, and retrieval, which is coupled with a compatible relational database schema that provides a declarative method for standardizing data and queries and allows users to directly state what information the database contains and what information the users want to extrapolate from it. As shown in Figure 2, the main structure of this database schema is designed and constructed as a comprehensive database containing seven tables that allows for the collection and storage of the following information from each and every study: (1) experiment identification, (2) sample identification (experimental model systems), (3) treatment information, (4) readout measurements defined by specific nomenclatures, (5) readout dictionary describing the meaning of these readout nomenclatures, (6) known relationships among the experimental/treatment paradigms, readouts, and their biological significances, and (7) target coefficients. Sample level information relating specific treatment paradigms and outcome measures (which we refer to as expert knowledge) was directly stored in a table of known relations where information was organized in the form of ordered triplets: test parameter, and how it is related to specific readout information/values. The collection of triplets (from individual experiments or summations from multiple experiments) we generated from the overall data set served as the “network building blocks” for Integrative Data Analysis, leading to the construction of a statistical network that graphically connects among the multitudes of treatment and readout nodes across the heterogeneous data source/types and depicts the direction of individual connections. Results were organized into a directed network graph showing statistically significant connections (e.g., GSPE treatment and bioavailability of specific phenolic metabolites), or connections based on prior knowledge (e.g., CGJ is a component of BDPP). Details on the construction of the three major components of our comprehensive database and analysis framework are presented in the Experimental Section.

Integrative Data Analysis and Interpretation

Using this framework, we conducted an integrated analysis of the multiscale BDPP data types encompassing all of our archived BDPP data sets. The collective BDPP data set we subjected to integrative analysis was generated from 47 independent preclinical experiments, involving either *in vitro* or *in vivo* model systems, with 19 treatment compounds, 138 readout measures, and a total of 1,769 experimental samples. These multiscale data types that were incorporated into the collective BDPP data set for integrative analysis are composed of data sets on the chemical composition of BDPP (Supplementary Figures S1 and S3A), the pharmacokinetics and biological availability of BDPP and BDPP components in plasma and in the brain (Supplementary Figures S2A–D and S3B), the effects of BDPP treatment on behavioral phenotypes relating to depression and cognition (Supplementary Figure S3C), the biological activities of BDPP and BDPP-derived phenolic metabolites on diverse cellular and molecular processes in *in vivo* or *in vitro* experimental paradigms (Supplementary Figure S3D,E), and toxicity profile and drug-like properties of select BDPP-derived phenolic compounds in *in vivo* or *in vitro* experimental paradigms (Supplementary Figure S3F,G).

Outcome of the Statistical Data Analysis

Our overall BDPP data set for integrative analysis contains 170 distinct statistical models, for a total of 924 model coefficients (including intercept terms), further summarized into 250 treatment-readout quantified associations. Statistical analysis at the sample level identified 50 nominally significant treatment–readout associations ($p < 0.01$). The collection of sample-level information provides the building blocks for the construction of an unbiased integrated network that connects BDPP administration and effects to a variety of biological/behavioral phenotypes. This integrated network construction was conducted by computer modeling using the *igraph* software (Version 1.0.1), following certain specified network construction rules. The resultant interpretive connection network was then annotated for clarity using the *Cytoscape* software (Version 3.4.0) and is presented in Figure 3. Outcomes from our integrative analysis connected BDPP treatment with multitudes of biological/behavioral phenotypes (Figure 3A). The resultant integrated graph is composed of a multitude of nodes, each representing a specific treatment paradigm, outcome measure, or expert knowledge. Nodes were then organized based on edges, which are relationships between nodes, with positive and negative connections between nodes depicted, respectively, with solid and dotted arrows. The unbiased organization of our overall data set resulted in an integrative network in which the overall biological complexity of BDPP is segregated and visualized in terms of CGJ, RSV, and GSPE, which are the three polyphenol-rich dietary preparations that constitute the BDPP mixture. Also depicted are multitudes of biologically available phenolic metabolites that are derived from CGJ, RSV, and GSPE as well as their connections to select biological activities. In Figure 3A, we also depict in the lower right-hand corner the effects of treating animals with quercetin (Q) on the generation of biologically available quercetin metabolites (Q-Gln and Me-Q-Gln) and the biological activities of quercetin treatment.

Interactions among the Overall Integrative Network Connecting BDPP Treatment with the Promotion of Psychological and Cognitive Resilience

Outcomes from our integrative analysis highlight connectivity between BDPP treatment and psychological resilience and cognitive resilience, which are shown, respectively, in Figure 3B and Figure 3C. In regard to depression/anxiety, current evidence suggests that elevated expression of the pro-inflammatory cytokine inter-leukin-6 (IL-6) in the periphery and downregulation of Ras-related C3 botulinum toxin substrate-1 (Rac-1) expression in the nucleus accumbens (NAc) in the brain both play key roles in desensitizing the brain reward system, thereby increasing sensitivity to stress-mediated depression/anxiety phenotypes.^{35–38} As highlighted by red arrows in Figure 3B, our integrative analysis revealed that IL-6 expression is significantly inhibited by select biologically available phenolic metabolites derived from BDPP treatment, particularly 3-(3'-hydroxyphenyl)propionic acid (3-HPP), dihydrocaffeic acid (DHCA), 3'-*O*-Me-epicatechin-5-*O*-glucuronide (Me-EC-Gln), and quercetin-3-*O*-glucuronide (Q-Gln). Moreover, Rac-1 expression is significantly induced by malvidin-glucoside, another BDPP metabolite. Based on outcomes from our integrative analysis showing the efficacy of DHCA in inhibiting IL-6 and malvidin-glucoside (M-Glc) in promoting Rac-1 expression, we have conducted an *in vivo* study to investigate the potential role of treatment with combined DHCA and M-Glc, for treating anxiety/depression. Consistent with our evidence (Figure 3B), we found that combined

dihydrocaffeic acid and malvidin-glucoside treatment significantly reduced development of depression/anxiety behavioral phenotypes in mice exposed to environmental stress following repeated social defeat.⁵⁰

Our integrative analysis also highlights connectivity between BDPP treatment and cognitive resilience (Figure 3C). Cognitive dysfunction has been mechanistically linked with reduced expression of phosphorylated cAMP-responsive element-binding protein (p-CREB), activity-regulated cytoskeleton-associated protein (Arc), and Fos proto-oncogene (c-Fos) in the hippo-campal region of the brain.^{39–43} As highlighted by red arrows in Figure 3C, our integrative analysis revealed that select BDPP-derived biologically available phenolic metabolites, particularly M-Glc, delphinidin-glucoside (D-glc), and Q-Gln, significantly promote p-CREB, Arc, and c-Fos expression.

Collectively, our integrative analysis highlights and allows for the systematic organization of connections between the benefits of BDPP treatment on psychological and cognitive resilience with the generation of select biologically available phenolic metabolites and effects of these metabolites in the modulation of IL-6 and Rac-1, p-CREB, Arc, and/or c-Fos in various experimental paradigms.

DISCUSSION

There is increasing interest in the development of select botanicals for the preservation of health and/or modulating multiple disease conditions.^{1–8} These “bioactive” botanical preparations are characterized by diverse and complex chemical compositions.^{17,18,25,44–47} Thus, the development of a given botanical preparation for eventual clinical applications needs to factor in multiple issues, including chemical composition characteristics of the botanical targeted for development, chemical consistency and reproducibility of the botanical preparation, dose-responsive biological activity and safety profiling of the botanical preparation, pharmacokinetics and biological availability of chemical constituents from the botanical preparation, and biological activities of these biologically available chemical constituents. Successful preclinical and clinical development of a given bioactive botanical will require a better understanding not only of each of these diverse issues but also of their interactions. Unfortunately, most botanical investigations typically address only a subset of these issues. Moreover, among the literature, investigations on a given botanical are typically presented in heterogeneous data types due to the usage of diverse preparations and dosages of the botanical, with testing being conducted in different experimental protocols and model systems limiting the ability to draw in larger inferences. We argue that systematic integrations of these heterogeneous data sources and types could significantly improve the approach to develop botanical supplements for use in preclinical and clinical studies.

In this study, we implement a comprehensive database to accommodate and standardize diverse data sets, including dietary dosage, metabolite concentration in the blood and brain, mRNA level of key genes, and behavior end points (e.g., stress and cognitive function) measured in cell cultures and animal models. Once established, the database at minimum serves three critical purposes. (1) High throughput quality control: Summary statistics for

every trait were automatically generated, and potential outliers and errors were highlighted for researchers to review/correct. Critically, the quality control step checks consistency and compatibility of data from different types of experiments. (2) Fast assessment of reproducibility/variation of the experimental results across laboratories, time, and experimental setups. (3) Data security, safety, and accessibility: Quality controlled data sets are stored in a centralized database protected by a firewall. The data sets are under systematic version controls, and changes/additions to the database are documented in a log file. The strict version control and change-tracking is essential in managing large numbers of data files accumulated over multiple years, and the iterative quality control process. The data sets are automatically backed up to archive and data vault periodically (e.g., every month) to guard against disk failure or data loss. The centralized database also allows easy and fast data query, where analysts can quickly retrieve data sets from multiple related experiments for integrated analysis.

We presented an integrative analysis framework with two key functions: (1) Estimation and testing association between traits measured in the same experiments. Given the study design (e.g., dose–response assay), the causal direction of the association can often be determined. (2) Construction of interaction networks. The relationships among traits that are not measured in the same experiments are inferred indirectly.

We developed a comprehensive database and analysis framework (see Experimental Section) to facilitate systematic collection of diverse, heterogeneous data types we have compiled from BDPP studies (see Supplementary Data Figures S1, 2, and 3). We applied our approaches to our heterogeneous collection of BDPP databases. Readouts from $N = 47$ experiments (both *in vivo* and *in vitro*) were standardized in the database and underwent careful quality control and correction. By integrating these experiments into a common framework, we created a visualized map of interactions among all of these factors and connected BDPP treatment with the promotion of psychological and cognitive resilience.

Outcomes from our integrative analysis highlight connections between the benefits of BDPP treatment with the generation of select biologically available phenolic metabolites that are effective in modulating known key cellular/molecular processes underlying depression/anxiety (e.g., IL-6 and Rac-1^{35–38}) and cognitive dysfunction (e.g., p-CREB, Arc, and/or c-Fos^{39–43}). Insights gathered from our integrative analysis also stimulated new investigations to test the mechanisms underlying the benefits of BDPP in promoting psychological and cognitive resilience. For example, our integrative analysis connected two BDPP-derived phenolic metabolites, DHCA and M-glc, with the promotion of psychological resilience by mechanistically inhibiting IL-6 expression in the periphery and inducing Rac-1 expression in the NAc (Figure 3B). Based on that, we explored the effects of combined DHCA/M-glc for treating anxiety and depression. Excitingly, we found that DHCA/M-glc treatment significantly reduced development of depression/anxiety behavioral phenotypes in mice that were subjected to repeated social defeat [Wang et al., submitted], which shed light on novel molecular crosstalking between peripheral and central pathophysiological cascades as novel therapeutic targets for treating depression.

The procedures for data validation, extraction, and modeling were coded and fully automated. This will allow researchers to update analysis results in a straightforward manner as new data will be collected, as well as to audit and revise previously obtained results, which will be fully reproducible *in silico*. We note that the biggest challenge that this study faced was the large degree of heterogeneity of the different data sets which we set to integrate into one single analysis. This was further exacerbated by the use of imprecise and often inconsistent nomenclature across laboratories or even between researchers in the same lab. While *ad hoc* analysis of each single experiment, often by the same people who performed the experiment, did not require further processing, standardization of formats, layouts, and terminology was an essential prerequisite for any kind of joint analysis. Even though the schema we converged upon could still be improved, we believe that we reached a satisfactory compromise between flexibility of the data format, to accommodate experiments of different aims and designs within a single database, and the rigidity required to perform more formal processing across different experiments.

A limitation of our currently constructed database and analysis framework is that the storage system is not optimized for the storage/retrieval of high throughput data sets that are generated using RNaseq, microarray, or other high throughput technologies. We are currently incorporating the use of a SQLite database engine^{48,49} while retaining the single-file benefits of the currently used spreadsheet to facilitate incorporation and retrieval of large high throughput data sets for integrative analysis in the future. Another limitation of our database and analysis framework is that we modeled for linear effects, but we did not characterize for nonlinear effects. This is because the sample size of each experiment is still relatively small and we do not have a sufficient degree of freedom to characterize nonlinear patterns. The ongoing research in our group continues to accumulate data, and will eventually achieve sufficiently large sample sizes for studying nonlinear patterns.

In sum, outcomes from our integrative analysis of our multi-scale, heterogeneous data types and data sets lead to a better understanding of how a complex polyphenol-rich preparation such as BDPP may mechanistically promote resilience against environmental stress-mediated psychological and cognitive dysfunctions, and provide a framework for continuation of pre-clinical and clinical development of BDPP for clinical applications. We realize that this model is in an early stage of development, and the conclusions drawn here are preliminary in nature. As more data are included in the database, the approach will become more powerful and useful in the future. The database and analysis framework is currently accessible only to internal users. We are working toward adapting our analytical framework and developing the appropriate rules, regulations, and guidelines to allow outside researchers to (i) access the database, (ii) query/retrieve data, and (iii) submit data and query data. We envision that our analytical framework can easily be adapted for integrative analysis toward the translational development of other complex botanical preparations by the research community.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported in part by Grant P50 AT008661-01 from the NCCIH and the ODS. G.M.P. holds a VA Career Scientist Award. We acknowledge that the contents of this study do not represent the views of the NCCIH, the ODS, the NIH, the U.S. Department of Veterans Affairs, or the United States Government. Funds for this research were also provided by the New Jersey Agricultural Experiment Station, NIFA HATCH 1005685/NJ 12158, and the New Use Agriculture and Natural Plant Products Program. K.H. is partially supported by the National Natural Science Foundation of China (Grants 21477087, 91643201) and by the Ministry of Science and Technology of China (Grant 2016YFC0206507). The study was also supported, in part, by the New Jersey Agricultural Experiment Station, NIFA HATCH 1005685/NJ 12158, and the New Use Agriculture and Natural Plant Products Program.

ABBREVIATIONS USED

3-HHA	3-hydroxyhippuric acid
3-HPA	3-hydroxyphenylacetic acid
3-HPP	3-(3'-hydroxyphenyl)propionic acid
4-HBA	4-hydroxybenzoic acid
5-HPV	5-(4'-hydroxyphenyl)valeric acid
4-HHA	4-hydroxyhippuric acid
Arc	activity-regulated cytoskeleton-associated protein
AGER	advanced glycosylation end product specific receptor
BDPP	bioactive dietary polyphenol preparation
C	catechin
c-Fos	Fos proto-oncogene
CGJ	Concord grape juice
CNTN1	β -catenin gene
Cy-glc	cyanidin-glucoside
CYP	cytochrome P450h
D-glc	delphinidin-glucoside
DHCA	dihydrocaffeic acid
diHPA	3,4-dihydroxyphenylacetic acid
EC	epicatechin
EC-Gln	epicatechin-5- <i>O</i> -glucuronide
Egr	early growth response protein
FA	ferulic acid

FK	foreign key
GLP-1	glucagon-like peptide 1
GOT	glutamic oxaloacetic transaminase
GSPE	grape seed polyphenol extract
Me-C-gln	3'- <i>O</i> -Me-catechin-5- <i>O</i> -glucuronide
HA	hippuric acid
IL-6	interleukin-6
hERG	human ether-à-go-go-related gene
JAMA	junction adhesion molecule A
LRP1	low density lipoprotein receptor-related protein 1
MCH	mean corpuscular hemoglobin
MCP-1	monocyte chemoattractant protein-1
M-glc	malvidin-glucoside
Me-Ec-Gln	3'- <i>O</i> -Me-epicatechin-5- <i>O</i> -glucuronide
Me-Q-gln	Me-quercetin- <i>O</i> -glu-curonide
<i>n</i>	number of samples
P70S6K	ribosomal protein S6 kinase beta-1
PAI-1	plasminogen activator inhibitor-1
p-CREB	phosphorylated cAMP-responsive element-binding protein
PECAM	platelet endothelial cell adhesion molecule
PERK1/2	phosphor-extracellular signal-regulated kinases 1/2
PK	primary key
Pt-glc	petunidin-glucoside
Q	quercetin
Q-gln	quercetin-3- <i>O</i> -glucuronide
Rac-1	Ras-related C3 botulinum toxin substrate-1
RSV	<i>all-trans</i> -resveratrol
RSV-gln	resvera-trol-3- <i>O</i> -glucuronide
SD	standard deviation

ZO1	zona occludens-1
VGAT	vesicular inhibitory amino acid transporter

References

1. Queen BL, Tollefsbol TO. Polyphenols and aging. *Curr Aging Sci.* 2010; 3(1):34–42. [PubMed: 20298168]
2. Abdal Dayem A, Choi HY, Yang GM, Kim K, Saha SK, Cho SG. The anti-cancer effect of polyphenols against breast cancer and cancer stem cells: molecular mechanisms. *Nutrients.* 2016; 8(9):581–617.
3. Chen AY, Chen YC. A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. *Food Chem.* 2013; 138:2099–2107. [PubMed: 23497863]
4. Khurana S, Venkataraman K, Hollingsworth A, Piche M, Tai TC. Polyphenols: benefits to the cardiovascular system in health and in aging. *Nutrients.* 2013; 5(10):3779–3821. [PubMed: 24077237]
5. Jiang F, Chang CWT, Dusting GJ. Cytoprotection by natural and synthetic polyphenols in the heart: novel mechanisms and perspectives. *Curr Pharm Des.* 2010; 16:4103–4112. [PubMed: 21247397]
6. Pasinetti GM. Novel role of red wine-derived polyphenols in the prevention of Alzheimer's disease dementia and brain pathology: experimental approaches and clinical implications. *Planta Med.* 2012; 78:1614–1619. [PubMed: 23023952]
7. Spagnuolo C, Napolitano M, Tedesco I, Moccia S, Milito A, Russo GL. Neuroprotective role of natural polyphenols. *Curr Top Med Chem.* 2016; 16(17):1943–1950. [PubMed: 26845551]
8. Pasinetti GM, Wang J, Ho L, Zhao W, Dubner L. Roles of resveratrol and other grape-derived polyphenols in Alzheimer's disease prevention and treatment. *Biochim Biophys Acta, Mol Basis Dis.* 2015; 1852(6):1202–1208.
9. Perez-Jimenez J, Neveu V, Vos F, Scalbert A. Identification of the 100 richest dietary sources of polyphenols: an application of the phenol-explorer database. *Eur J Clin Nutr.* 2010; 64:S112–S120. [PubMed: 21045839]
10. Joseph SV, Edirisinghe I, Burton-Freeman BM. Fruit Polyphenols: A review of anti-inflammatory effects in humans. *Crit Rev Food Sci Nutr.* 2016; 56:419–444. [PubMed: 25616409]
11. Umeno A, Horie M, Murotomi K, Nakajima Y, Yoshida Y. Antioxidative and antidiabetic effects of natural polyphenols and isoflavones. *Molecules.* 2016; 21(6):708–722.
12. Fraga CG, Galleano M, Verstraeten SV, Oteiza PI. Basic biochemical mechanisms behind the health benefits of polyphenols. *Mol Aspects Med.* 2010; 31:435–445. [PubMed: 20854840]
13. Ho L, Ferruzzi MG, Janle EM, Wang J, Gong B, Chen TY, Lobo J, Cooper B, Wu QL, Talcott ST, Percival SS, Simon JE, Pasinetti GM. Identification of brain-targeted bioactive dietary quercetin-3-O-glucuronide as a novel intervention for Alzheimer's disease. *FASEB J.* 2013; 27:769–781. [PubMed: 23097297]
14. Huang HP, Ou TT, Wang CJ. Mulberry and its Bioactive Compounds, the Chemoprevention Effects and Molecular Mechanisms In Vitro and In Vivo. *J Tradit Complement Med.* 2013; 3:7–15. [PubMed: 24716151]
15. Jung KJ, Wallig MA, Singletary KW. Purple grape juice inhibits 7,12-dimethylbenz[a]anthracene (DMBA)-induced rat mammary tumorigenesis and in vivo DMBA-DNA adduct formation. *Cancer Lett.* 2006; 233:279–288. [PubMed: 15878797]
16. Shanmuganayagam D, Warner TF, Krueger CG, Reed JD, Folts JD. Concord grape juice attenuates platelet aggregation, serum cholesterol and development of atheroma in hypercholesterolemic rabbits. *Atherosclerosis.* 2007; 190:135–142. [PubMed: 16780846]
17. Villani TS, Reichert W, Ferruzzi MG, Pasinetti GM, Simon JE, Wu Q. Chemical investigation of commercial grape seed derived products to assess quality and detect adulteration. *Food Chem.* 2015; 170:271–280. [PubMed: 25306345]
18. Xu Y, Simon JE, Welch C, Wightman JD, Ferruzzi MG, Ho L, Pasinetti GM, Wu Q. Survey of polyphenol constituents in grapes and grape-derived products. *J Agric Food Chem.* 2011; 59(19): 10586–10593. [PubMed: 21879745]

19. Dubner L, Wang J, Ho L, Ward L, Pasinetti GM. Recommendations for development of new standardized forms of cocoa breeds and cocoa extract processing for the prevention of Alzheimer's disease: role of cocoa in promotion of cognitive resilience and healthy brain aging. *J Alzheimer's Dis.* 2015; 48(4):879–989. [PubMed: 26402120]
20. Manach C, Scalbert A, Morand C, Remesy C, Jimenez L. Polyphenols: food sources and bioavailability. *Am J Clin Nutr.* 2004; 79(5):727–747. [PubMed: 15113710]
21. Calani L, Dall'Asta M, Derlindati E, Scazzina F, Bruni R, Del Rio D. Colonic Metabolism of Polyphenols From Coffee, Green Tea, and Hazelnut Skins. *J Clin Gastroenterol.* 2012; 46:S95–S99. [PubMed: 22955368]
22. Hribar U, Ulrich NP. The metabolism of anthocyanins. *Curr Drug Metab.* 2014; 15:3–13. [PubMed: 24329109]
23. Lambert JD, Sang S, Lu AYH, Yang CS. Metabolism of dietary polyphenols and possible interactions with drugs. *Curr Drug Metab.* 2007; 8:499–507. [PubMed: 17584021]
24. Monagas M, Urpi-Sarda M, Sanchez-Patan F, Llorach R, Garrido I, Gomez-Cordoves C, Andres-Lacueva C, Bartolome B. Insights into the metabolism and microbial biotransformation of dietary flavan-3-ols and the bioactivity of their metabolites. *Food Funct.* 2010; 1:233–253. [PubMed: 21776473]
25. Wang J, Tang C, Ferruzzi MG, Gong B, Song BJ, Janle EM, Chen TY, Cooper B, Varghese M, Cheng A, Freire D, Bilski A, Roman J, Nguyen T, Ho L, Talcott ST, Simon JE, Wu Q, Pasinetti GM. Role of standardized grape polyphenol preparation as a novel treatment to improve synaptic plasticity through attenuation of feature of metabolic syndrome in a mouse model. *Mol Nutr Food Res.* 2013; 57(12):2091–2102. [PubMed: 23963661]
26. Wang J, Bi W, Cheng A, Freire D, Vempati P, Zhao W, Gong B, Janle EM, Chen TY, Ferruzzi MG, Schmeidler J, Ho L, Pasinetti GM. Targeting multiple pathogenic mechanisms with polyphenols for the treatment of Alzheimer's disease – experimental approach and therapeutic implications. *Front Aging Neurosci.* 2014; 6:42. [PubMed: 24672477]
27. Zhao W, Wang J, Bi W, Ferruzzi M, Yemul S, Freire D, Mazzola P, Ho L, Dubner L, Pasinetti GM. Novel application of brain-targeting polyphenol compounds in sleep deprivation-induced cognitive dysfunction. *Neurochem Int.* 2015; 89:191–197. [PubMed: 26235983]
28. Ferruzzi MG, Lobo JK, Janle EM, Cooper B, Simon JE, Wu QL, Welch C, Ho L, Weaver C, Pasinetti GM. Bioavailability of gallic acid and catechins from grape seed polyphenol extract is improved by repeated dosing in rats: implications for treatment in Alzheimer's disease. *J Alzheimer's Dis.* 2009; 18(1):113–124. [PubMed: 19625746]
29. Wang J, Ferruzzi MG, Ho L, Blount J, Janle EM, Gong B, Pan Y, Gowda GA, Raftery D, Arrieta-Cruz I, Sharma V, Cooper B, Lobo J, Simon JE, Zhang C, Cheng A, Qian X, Ono K, Teplow DB, Pavlides C, Dixon RA, Pasinetti GM. Brain-targeted proanthocyanidin metabolites for Alzheimer's disease treatment. *J Neurosci.* 2012; 32(15):5144–5150. [PubMed: 22496560]
30. Vingtdeux V, Giliberto L, Zhao H, Chandakkar P, Wu Q, Simon JE, Janle EM, Lobo J, Ferruzzi MG, Davies P, Marambaud P. AMP-activated protein kinase signaling activation by resveratrol modulates amyloid-beta metabolism. *J Biol Chem.* 2010; 285(12):9100–9113. [PubMed: 20080969]
31. Wang D, Ho L, Faith J, Ono K, Janle EM, Lachcik PJ, Cooper BR, Jannasch AH, D'Arcy BR, Williams BA, Ferruzzi MG, Levine S, Zhao W, Dubner L, Pasinetti GM. Role of intestinal microbiota in the generation of polyphenol-derived phenolic acid mediated attenuation of Alzheimer's disease β -amyloid oligomerization. *Mol Nutr Food Res.* 2015; 59(6):1025–1040. [PubMed: 25689033]
32. Wang J, Santa-Maria I, Ho L, Ksiezak-Reding H, Ono K, Teplow DB, Pasinetti GM. Grape-derived polyphenols attenuate tau neuropathology in a mouse model of Alzheimer's disease. *J Alzheimer's Dis.* 2010; 22(2):653–661. [PubMed: 20858961]
33. Pflieger CM, Wang J, Friedman L, Vittorino R, Conley LM, Ho L, Fivecoat HC, Pasinetti GM. Grape-seed polyphenolic extract improves the eye phenotype in a *Drosophila* model of tauopathy. *Int J Alzheimer's Dis.* 2010; 2010:576357. [PubMed: 20871666]
34. Liu P, Kemper LJ, Wang J, Zahs KR, Ashe KH, Pasinetti GM. Grape seed polyphenolic extract specifically decreases $a\beta^{*56}$ in the brains of Tg2576 mice. *J Alzheimer's Dis.* 2011; 26(4):657–666. [PubMed: 21743132]

35. Köhler CA, Freitas TH, Maes M, de Andrade NQ, Liu CS, Fernandes BS, Stubbs B, Solmi M, Veronese N, Herrmann N, Raison CL, Miller BJ, Lanctôt KL, Carvalho AF. Peripheral cytokine and chemokine alterations in depression: a meta-analysis of 82 studies. *Acta Psychiatr Scand*. 2017; 135(5):373–387. [PubMed: 28122130]
36. Fonseka TM, McIntyre RS, Soczynska JK, Kennedy SH. Novel investigational drug targeting IL-6 signaling for the treatment of depression. *Expert Opin Invest Drugs*. 2015; 24(4):459–475.
37. Golden SA, Christoffel DJ, Heshmati M, Hodes GE, Magida J, Davis K, Cahill ME, Dias C, Ribeiro E, Ables JL, Kennedy PJ, Robison AJ, Gonzalez-Maeso J, Neve RL, Turecki G, Ghose S, Tamminga CA, Russo SJ. Epigenetic regulation of Rac1 induces synaptic remodeling in stress disorders and depression. *Nat Med*. 2013; 19(3):337–344. [PubMed: 23416703]
38. Lenz JD, Lobo MK. Optogenetic insights into striatal function and behavior. *Behav Brain Res*. 2013; 255:44–54. [PubMed: 23628212]
39. Kempf SJ, Casciati A, Buratovic S, Janik D, von Toerne C, Ueffing M, Neff F, Moertl S, Stenerlöv B, Saran A, Atkinson MJ, Eriksson P, Pazzaglia S, Tapio S. The cognitive defects of neonatally irradiated mice are accompanied by changed synaptic plasticity, adult neurogenesis and neuroinflammation. *Mol Neurodegener*. 2014; 9:57. [PubMed: 25515237]
40. Zhang XQ, Mu JW, Wang HB, Jolkkonen J, Liu TT, Xiao T, Zhao M, Zhang CD, Zhao CS. Increased protein expression levels of pCREB, BDNF and SDF-1/CXCR4 in the hippocampus may be associated with enhanced neurogenesis induced by environmental enrichment. *Mol Med Rep*. 2016; 14(3):2231–2237. [PubMed: 27432087]
41. Lin Y, Lu X, Dong J, He X, Yan T, Liang H, Sui M, Zheng X, Liu H, Zhao J, Lu X. Involuntary, Forced and Voluntary Exercises Equally Attenuate Neurocognitive Deficits in Vascular Dementia by the BDNF-pCREB Mediated Pathway. *Neurochem Res*. 2015; 40(9):1839–1848. [PubMed: 26240057]
42. Barrientos RM, Frank MG, Watkins LR, Maier SF. Aging-related changes in neuroimmune-endocrine function: implications for hippocampal-dependent cognition. *Horm Behav*. 2012; 62(3):219–227. [PubMed: 22370243]
43. Zhong T, Ren F, Huang CS, Zou WY, Yang Y, Pan YD, Sun B, Wang E, Guo QL. Swimming exercise ameliorates neurocognitive impairment induced by neonatal exposure to isoflurane and enhances hippocampal histone acetylation in mice. *Neuroscience*. 2016; 316:378–388. [PubMed: 26748054]
44. Ho L, Chen LH, Wang J, Zhao W, Talcott ST, Ono K, Teplow D, Humala N, Cheng A, Percival SS, Ferruzzi M, Janle E, Dickstein DL, Pasinetti GM. Heterogeneity in red wine polyphenolic contents differentially influences Alzheimer's disease-type neuropathology and cognitive deterioration. *J Alzheimer's Dis*. 2009; 16(1):59–72. [PubMed: 19158422]
45. Tao W, Zhou Z, Zhao B, Wei T. Simultaneous determination of eight catechins and four theaflavins in green, black and oolong tea using new HPLC-MS-MS method. *J Pharm Biomed Anal*. 2016; 131:140–145. [PubMed: 27589031]
46. Kusznierevicz B, Piekarska A, Mrugalska B, Konieczka P, Namie nik J, Bartoszek A. Phenolic composition and antioxidant properties of Polish blue-berried honeysuckle genotypes by HPLC-DAD-MS, HPLC postcolumn derivatization with ABTS or FC, and TLC with DPPH visualization. *J Agric Food Chem*. 2012; 60(7):1755–1763. [PubMed: 22264130]
47. Pawar RS, Handy SM, Grundel E, Cheng R, Shyong N. Assessment of the authenticity of herbal dietary supplements: comparison of chemical and DNA barcoding methods. *Planta Med*. 2017; doi: 10.1055/s-0043-107881
48. Zeng H, Qiu C, Cui Q. Drug-path: a database for drug-induced pathways. *Database*. 2015; 2015:bav061. [PubMed: 26130661]
49. Zhu Y, Davis S, Stephens R, Meltzer PS, Chen Y. CEOMetadb: powerful alternative search engine for the Gene Expression Omnibus. *Bioinformatics*. 2008; 24(23):2798–2800. [PubMed: 18842599]
50. Wang J, Hodes GS, Zhang H, Zhang S, Zhao W, Golden SA, Hi W, Menard C, Kana V, Leboeuf M, Xie M, Bregman D, Pfau M, Shanahan M, Esteban-Fernández A, Yemul S, Sharma A, Ho L, Dixon R, Nerad M, Han MH, Russo SJ, Pasinetti GM. Epigenetic modulation of inflammation and synaptic plasticity promotes resilience against stress disorder and depression. *Nat Commun*. 2018; doi: 10.1038/s41467-017-02794-5

Figure 1.

Schematic representation of our overall strategy toward establishing a comprehensive database and analysis framework for collecting and analyzing heterogeneous data types we have gathered using diverse experimental model systems and protocols. The framework consists of three phases. The first phase is Data Collection and Collation, which includes organization of the overall database (DB schema) gathered from the heterogeneous data types, standardization of nomenclature from all data entries, and quality control of the entries. Information generated is then subjected to Sample Level Data Analysis, which is a second phase analysis by the framework that includes expert knowledge of how individual studies in the overall database were designed and executed and readout values regarding polyphenol bioavailability, bioactivity, behavioral phenotypes, and/or safety/toxicity gathered in each study, to be followed by statistical analysis of outcome measures from individual studies. The information gathered is then further assessed by Integrative Data Analysis and Interpretation, which statistically integrates outcomes from all studies into the database. Integrated analysis outcomes are then statistically ordered and ranked and presented in a table form to highlight the most robust interactions within components of the overall database.

Figure 2.

Schematic representation of the core structure of our database schema. Our collection of heterogeneous and multiscale data types relating to BDPP was organized into a relational database. The graphic outlines the finalized schema. The collective database was organized into seven main tables, which are linked to each other through unique identifiers such as defined experiments, samples, and readouts. Based on this schema, the overall database is designed and constructed as a comprehensive Excel file composed of seven tables. This collective database includes (1) “hard” experimental data that are collected primarily into the “readout_measurements” table; (2) expert knowledge, such as the biological significance of individual experimental observations, that is collected into the “known_relations” table; and (3) metadata which give information about relationships between specific experimental observations with other data that are collected into the “readout_dictionary” and the “target_coefficients” tables.

Figure 3.

Construction of an integrated network of natural compound activities reveals mechanistic pathways by which BDPP might help promote resilience against depression/anxiety and cognitive dysfunctions. (A) Derivation of an overall integrative network connecting BDPP with multitudes of biological/behavioral phenotypes. The thick gray arrows stemming from BDPP indicate the three BDPP components: GSPE, RSV, and CGJ. The network graph connecting individual BDPP components is composed of a multitude of nodes, each representing a molecule. Nodes were organized with connections, standard for prior knowledge or significant statistical associations (p -value ≤ 0.01). Thin solid and dotted arrows depict edges and the direction of individual connections between nodes, with positive and negative connections between nodes depicted, respectively, by solid and dotted arrows. (B, C) Nodes connecting BDPP treatment with the promotion of psychological and cognitive resilience. The two schematics in panels B and C are identical to the one shown in panel A, but highlight the connectivity between BDPP treatment and psychological resilience (B) and cognitive resilience (C). In panel B, red arrows identify connections between the BDPP treatment with the generation of select biologically available phenolic metabolites and the effects of these metabolites in the modulation of IL-6 and Rac-1 that play key roles in psychological resilience. In panel C, red arrows identify connections between the BDPP

treatment with the generation of select biologically available phenolic metabolites and effects of these metabolites in the modulation of p-CREB, Arc, and c-Fos that play key roles in cognitive resilience.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript