Universidad Zaragoza
1542

Master Thesis

Identification of disease-related targets for red wine polyphenols

Author:
María Alejandra Sierra Aguilera

Supervisors:
Juan Fernández-Recio
María Milagros Medina Trullenque

Master of Biophysics and Quantitative biotechnology
January 2023

Instituto Universitario de Investigación Biocomputación y Física de Sistemas Complejos Universidad Zaragoza

Facultad de Ciencias Universidad Zaragoza

Quantitative Biotechnology

Table of Contents

1. INTRODUCTION

Polyphenols are phytochemicals that are present in plant-based food. According to Phenol-Explorer database (http://phenol-explorer.eu/), red wine is the food source with more variate of polyphenols, and different studies supports that these phenolic compounds shows beneficial effects like decreasing the risk of cardiovascular diseases, beneficial effects on lipid metabolism and anti-cancer properties.

Despite the fact that there is comprehensive information concerning the biological functions of polyphenols, evidence to support the clear effects on human health remain weak. (Tresserra-Rimbau et al., 2018). Their role in modulating the risk of diseases is difficult to demonstrate due to their more than 500 hundred different structures, which means different biological actions. Additional limitations are: the complexity of estimating the polyphenol content in food, and how it changes in case on how we cook them or ingest them, the interaction with the human body metabolism and the gut microbiome, the variety of experimental settings and limited scope of studies that explore the molecular effect.

In this context, computational strategies, including virtual screening, pharmacology network, and molecular docking, are a strong tool to identify potential targets of polyphenols, get more insights about their molecular mechanisms and gain better understanding of the therapeutic effects exerted of dietary polyphenols.

In the following study, we describe the development of a new comprehensive database of dietary polyphenols, initially focused on red wine, as a framework to capture the molecular interactions between polyphenols and diseases, genes, pathways, gene ontology and proteins, to propose directions for future research in the field of drug target identification and drug repurposing.

2. OBJECTIVES

2.1 General objective

- Develop a database of polyphenols in red wine and identify targets

2.2 Specific objectives

- Identify diseases related to polyphenols in red wine
- Identify genes related to polyphenols in red wine
- Identify proteins related to polyphenols in red wine
- Apply in one example the database

3. CONTEXT AND BACKGROUND

3.1 Polyphenols

Polyphenols are soluble natural secondary metabolites, resulting from two biosynthetic pathways: phenylpropanoid and/or the poliacetate -polyketide- pathway (Vong et al., 2022), presenting one or more hydroxylated aromatic ring -phenyl- moiety (Quideau et al., 2011), and with molecular masses up to 30,000 Daltons (Lacroix et al., n.d.).

These phenolic compounds are present in more than 400 food sources and beverages, like fruits, vegetables, whole grains, seeds, coffee, herbs, spices, green tea, and especially red wine, with a rich variety of phenolic compounds (do Valle et al., 2021; Neveu et al., n.d.).

3.2 Polyphenol classes and sub-classes

Phenol-Explorer database (version 3.6 http://phenol-explorer.eu/) have profiled over 500 polyphenols in more than 400 foods and have documented a high diversity of polyphenols to which humans are exposed through their diet. This diversity is classified based on the chemical structure features, in general into two main categories: flavonoids and non-flavonoids (Visioli et al., 2020). However, Phenol-Explorer classify them in 6 classes and at least 28 subclasses that are describe in the following section (Neveu et al., n.d.; Rothwell et al., 2013):

3.2.1 Flavonoids

The basic structure of this group has a skeleton with 15 carbon atoms in three rings labeled with letters A, B, and C in blue (see Figure 1). The oxygen atom is numbered as the first position, and the remaining carbon atoms are numbered from C2 to C10, in rings A and C, where C is the heterocyclic ring. Finally, Ring B is numbered from C1` to C6` (Panche et al., 2016; Singla et al., 2019; Visioli et al., 2020).

The structural diversity of this group depends on the carbon of the C ring on which the B ring is attached and the degree of unsaturation, oxidation, hydroxylation, prenylation, alkalization or glycosylation reactions of the C ring. And they are classified in 9 following subclasses (Panche et al., 2016; Singla et al., 2019; Visioli et al., 2020).

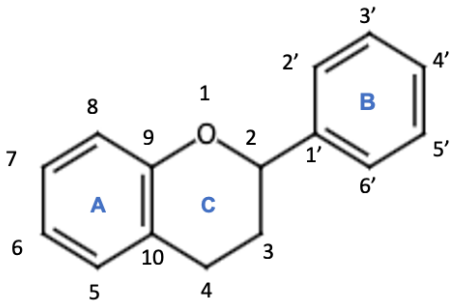


Figure 1: Basic structure of Flavonoids (SwissADME http://www.swissadme.ch/ was used to draw the structure).

3.2.1.1 Anthocyanins

They are the glycosylated form of anthocyanidins. They have two double bonds in their heterocyclic rings. Anthocyanins are characterized by the acylation or methylation patterns on rings A and B, and the nature and number of bonded monosaccharides e.g. glucose, galactose, and arabinose to their structure (Panche et al., 2016; Singla et al., 2019). The main anthocyanidins found in red wine are malvidin, petunidin, peonidin, delphinidin and cyanidin (Visioli et al., 2020).

3.2.1.2 Chalcones

Also named open-chain flavonoids. They are characterised by the absence of ring C (Panche et al., 2016).

3.2.1.3 Dihydrochalcones

One of the hydrogens of the methyl group is replaced by a benzyl group. They are compounds derived from 1,3-diphenylpropan-1-one (*Flavonoid Nomenclature*, n.d.).

3.2.1.4 Flavanols

Also called flavan-3-ols, because of the presence of a hydroxyl group in C3 position, which gives two chiral centers to this sub group of polyphenols. They contain a saturated heterocyclic ring, with no double bond between C2 and C3, which makes them benzopyrans (Visioli et al., 2020). Flavanols in red wine include aglycons, such as catechins and epicatechins. Their two chiral

centers make them diastereoisomers, where catechin isomer shows the trans configuration of: (+)- catechin and (–)- catechin, and epicatechin has the cis configuration of (+)- epicatechin and (–)-epicatechin stereoisomers (Panche et al., 2016; Singla et al., 2019). Catechins usually occur as aglycones, or esterified with gallic acid that can form polymers, named proanthocyanins, like procyanidin dimers B1 to B4, and trimers such as procyanidin C1 (Visioli et al., 2020).

3.2.1.5 Flavanones

In this group, unlike flavonols and flavones, the heterocyclic ring (ring C) in flavanones has a saturated three-carbon chain without a hydroxyl group at the C3 position. Therefore, the double bond between positions 2 and 3 is saturated. Flavanones have a large number of substituted derivatives for example prenylated flavanones, furanoflavanones, benzylated flavanones, pyranoflavanones. One example of this group found in red wine is naringenin (Panche et al., 2016; Singla et al., 2019; Visioli et al., 2020).

3.2.1.6 Dihydroflavonols

Are the 3-hydroxy derivatives of flavanones (Panche et al., 2016).

3.2.1.7 Flavones

They are characterized by the absence in C3·position of a hydroxyl group, the double bond between the C2 and C3 atoms, a carbonyl group in C4 position, and a ring B attached to the heterocyclic ring at the C2 position. Apigenin is part of this group present in red wine (Panche et al., 2016; Singla et al., 2019).

3.2.1.8 Flavonols

Also named 3-hydroxyflavones, because they have a hydroxyl group in the C3 position. They have a ketone group. Two of the most researched are kaempferol, myricetin and quercetin (Panche et al., 2016; Singla et al., 2019; Visioli et al., 2020).

3.2.1.9 Isoflavones

An oxygen atom is placed on the C4 position, and ring B is attached to the heterocyclic ring in C3 position (instead of C2 as in other classes) (Singla et al., 2019).

3.2.2 Lignans

Derived from derivatives of cinnamic acid. They belong to a nonflavonoid class with two propylbenzene units (C6-C3) linked together between the β-position in C8 of the propane side chains. The C9 and C9` positions are substituted with different molecules, resulting in a wide range of different structural forms (“Style and Usage for Organic Chemistry,” 2009).

3.2.3 Non-phenolic metabolites of polyphenols

This group includes 4 polyphenols: 4-Ethylbenzoic acid, Glycine, 1,3,5-Trimethoxybenzene and Vanilloylglycine (Neveu et al., n.d.).

3.2.4 Other polyphenols

This class includes the following sub-class: Alkylmethoxyphenols, Alkylphenols, Curcuminoids, Furanocoumarins, Hydroxybenzaldehydes, Hydroxybenzoketones, Hydroxycinnamaldehydes, Hydroxycoumarins, Hydroxyphenylpropenes, Methoxyphenols, Naphtoquinones, Phenolic terpenes, Tyrosols and Other polyphenols (Neveu et al., n.d.).

3.2.5 Phenolic acids

A class with predominantly benzoic acid and cinnamic acid derivatives. They are commonly found in conjugation with other polyphenols, glucose, quinic acid, or structural components of the original plant. There are two distinguishing parent skeletons of phenolic acids: hydroxycinnamic and hydroxybenzoic acids. But there are also Hydroxyphenylacetic acids, Hydroxyphenylpropanoic acids and Hydroxyphenylpentanoic acids (Singla et al., 2019).

3.2.6 Stilbenes

Nonflavonoids characterized by two phenyl moieties linked together by a two-carbon methylene group. In stilbenes, the m-positions in ring A are usually substituted by two hydroxyl groups, while various positions in ring B may be substituted by hydroxyl and methoxyl groups. Stilbenes exist in both isomeric forms (cis and trans configurations). One of the most well-known and researched stilbenes is resveratrol (Singla et al., 2019).

3.3 Polyphenols in red wine

Red wine is a traditionally alcoholic beverage obtained by fermentation of grape must, with approximately 78% water, and a complex group of bioactive components such as sugars, lipids, soluble, proteins, vitamins, esters, ketones, organic acids, and polyphenols (Fernandes et al., 2017; Markoski et al., 2016).

The amount of phenolic compounds in wine depend on different factors like grape variety, the winemaking process, type of yeast that participates in the fermentation, and whether grape solids are present in the maceration process (Baiano et al., 2009; Visioli et al., 2020). However, is known that when is compared red and white wine, the red wine is much richer in polyphenols (Markoski et al., 2016). In general, red wine is one of the 100 foods with the highest polyphenol content with 101 mg per 100 ml (Pérez-Jiménez et al., 2010) and is also the food with the largest variety of phenolic compounds according to Phenol-Explorer database.

Red wine polyphenols includes anthocyanins, flavanols, flavanones, flavones, flavonols, phenolic acids, and stilbenes. (Diker & Kutluay, 2021). Where Flavonoids account for over 60% and Catechin and epicatechin are usually the most important flavonols in grape skin and seed. Anthocyanins are responsible for the red color of wines and are extracted from grape skin, and the most commonly found are delphinidin-3-glucoside, cyanidin-3-glucoside, and malvidin-3-glucoside. And from the group of Stilbene resveratrol, that is present in skin and seeds, it has been more studied polyphenol, even though, its concentration in wine is lower than others. (Markoski et al., 2016)

3.4 Red Wine Polyphenols and Health benefits

The moderated consumption of red wine has been correlated to the decrease of the risk of cardiovascular diseases, type 2 diabetes mellitus, and disorders with inflammation component like neurodegenerative diseases. Interestingly, the beneficial effects of consuming wine have been attributed, at least in part, to the biological effects of their polyphenol content. (Tresserra-Rimbau et al., 2014).

Recent studies have analyzed the interaction between red wine polyphenols and diseases like cancer, diabetes, cardiovascular and neurodegenerative diseases, Alzheimer, atherosclerosis, where has been identify their anti-inflammatory and vasorelaxing activity (do Valle et al., 2021; Lucarini et al., 2021). Furthermore, consumers of alcohol that include moderate wine intake have reduced mortality compared with those that do not consume wine, due to a lower incidence of coronary heart disease and light wine intake could protect against nonalcoholic liver disease. There are studies that shows that these benefits could be the basis of the French paradox (Rothwell et al., 2013).

Despite many health benefits, there is still a lot of discussion about the real properties of its components and its actions on cells and molecular interactions. Most of the epidemiological studies do not specifically detail the type of consumed alcohol, and they include red wine together with all the rest of alcoholic beverages. Thus, a large part of these issues permeate the fine line between the amount of alcohol that causes problems to organic systems and the amount that could be beneficial for the health, and because the benefits have been identified only in vitro or pre-clinical, it is important to identify mechanisms of interaction between polyphenols and proteins in our body. Considering the complexity and diversity in structure and the chemical properties of dietary polyphenols and their metabolites in gut microbiome, it is indispensable to complement existing data with computational tool as molecular pharmacology and network pharmacology (Visioli et al., 2020).

3.5 Network pharmacology

It is a discipline that combines the functions of compounds, disease, targets and biological signaling pathways based on computational analysis. It is suitable to identify potential polyphenol

targets, helping with the challenge of the high variety of structures, by combining the information of different studies (Xinyi et al., 2022).

To create a network pharmacology, it is important to use proper data sources and searching engines. For this study the tools that have been used are the following.

PUBCHEM (version V1.7.2.b - 2022Feb)

PubChem (https://pubchem.ncbi.nlm.nih.gov/) is an open chemistry database at the National Institutes of Health (NIH) launched in 2004. The scientific data comes from 892 organizations, and it has more than 113,803,291 unique chemical structures (Kim et al., 2022).

LENS.ORG (Version 8.7.1)

Lens (https://www.lens.org/) Is an aggregator of metadata, with for four primary functions: Scholarly Works index (over 225 million scholarly works), patents index (127 million global patent records), patseq (370 million patent sequences) where there are biological sequences disclosed in patent literature and collections, which is management tool to track, monitor, and analyze a collection of works or a collection of patents (Jefferson et al., n.d.).

CHEBI (Release 216)

Chemical Entities of Biological Interest (ChEBI) https://www.ebi.ac.uk/chebi/ is a free dictionary of small molecular entities (do not include directly encoded by the genome molecules, therefore nucleic acids, proteins and peptides are excluded). In this database, it is possible to find any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, and others, identifiable as a separately distinguishable entity. It can be natural or synthetic products and part of molecular entities (like substituent groups or atoms).

Phenol-Explorer (version 3.6)

This is a database of polyphenol content in foods (http://phenol-explorer.eu/). Contains more than 35,000 content values for 500 different polyphenols in over 400 foods. These data are derived from scientific publications and It is possible to have information about metabolism, effects of food processing and cooking. It is developed by INRA (French National Research Institute for Agriculture, Food and Environment) in collaboration with AFSSA (French Agency for Food, Environmental and Occupational Health & Safety), the University of Alberta, the University of

Barcelona, IARC (International Agency for Research on Cancer of the World Health Organization), And In Siliflo (Rothwell et al., 2013).

CTD (January 2023, update):

CTD (http://ctdbase.org/) is database released on 2004, which provides manually curated information about chemical–gene/protein interactions, chemical–disease and gene–disease relationships. There are more than 51,500,000 relationships that are integrated with functional and pathway data to aid in development of hypotheses about the mechanisms underlying environmentally influenced diseases (Davis et al., 2022a).

STITCH: (version 5.0)

STITCH (http://stitch.embl.de/) is a database of known and predicted interactions between chemicals and proteins. Currently covers more than 9,500,000 proteins from 2031 organisms. The interactions include direct (physical) and indirect (functional) associations; The source of information are High-throughput Lab Experiments, Genomic Context Predictions, (Conserved) Co-Expression, Automated Textmining and Previous Knowledge in Databases (Szklarczyk et al., 2016).

STRING (version 11.5)

The “Search Tool for the Retrieval of Interacting Genes/Proteins” (STRING) (http://string-db.org/) database gives protein–protein association evidence. The associations are imported from other data basses, and predicts associations called “de novo”. It includes a confidence score to each prediction, that derive by compare the performance of the prediction against common references set of trusted to true associations (von Mering et al., 2005).

DAVID (version 2021)

The Database for Annotation, Visualization and Integrated Discovery (DAVID), (https://david.ncifcrf.gov/home.jsp) provides a set of functional annotation tools to understand the biological meaning of large lists of genes. It can be used to identify enriched biological themes, GO terms, discover enriched functional-related gene groups, cluster redundant annotation terms, visualize genes on BioCarta & KEGG pathway maps, and list interacting proteins, among other applications.

KEGG (version 2023)

Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.kegg.jp) is a manually curated database resource integrating various biological objects categorized into systems, genomic, chemical and health information. It helps to understand high-level functions and utilities of the biological system, like the cell, organism and ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (Kanehisa et al., 2023).

REACTOME

The Reactome Knowledgebase (https://reactome.org), is an open access, manually curated and peer-reviewed pathway database. Founded in 2003, it provides molecular details of signal transduction, transport, DNA replication, metabolism and other cellular processes as an ordered network of molecular transformations in a single consistent data model (Jassal et al., 2020).

CYTOSCAPE

CYTOSCAPE (https://cytoscape.org/) Is an open platform with many plugins for visualization options and the network analysis. There are multiple layers of information including large-scale, genome-wise experiments, and protein function annotations (Otasek et al., 2019). The plug in CytoHubba is a useful way for presenting biological data including protein-protein interactions (PPI), gene regulations, cellular pathways, and signal transductions. It is possible to measure nodes by their network features to infer their importance in the network, and it can help us to identify central elements of biological networks. It provides 11 topological analysis methods including Degree, Edge Percolated Component, Maximum Neighborhood Component, Density of Maximum Neighborhood Component, Maximal Clique Centrality and six centralities (Bottleneck, EcCentricity, Closeness, Radiality, Betweenness, and Stress) based on shortest paths (Chin et al., 2014).

4. METHODS

4.1 Building the database of polyphenols content in red wine

4.1.1 Phenol-Explorer

To generate the list of polyphenols the search was conducted through the Phenol-Explorer database version 3.6 (www.phenol-explorer.eu) (Neveu et al., n.d.) where it was retrieved 501 polyphenols. To identify the ones, present in red wine it was used the key word: Wine [red], resulting in 123 polyphenols. After the exploration and cleaning of the data it was filter out 19 items: six compounds that represents a general group of polyphenols (02 mers 03 mers 04-06 mers 07-10 mers Polymers (>10 mers) Polyphenols, total) and not specific names and 13 duplicates, where it was reported data for the same polyphenol with two different analyzing methods -Chromatography and Chromatography after hydrolysis-. For the analysis in this study, it was selected one method for each compound having the final list of 104 polyphenols in Phenol-Explorer related to red wine.

4.1.2 Research Literature

This list obtained by Phenol-Explorer, was enriched with literature search, performed on “Lens” (https://www.lens.org/), using key words: “red wine” AND “polyphenol content”. The research was set up in time range 2000-2022, evaluating that was used an appropriate analytical method, like Chromatography. HPLC, Chromatography after hydrolysis, where excluded: UV spectrophotometry, vanillin-HCl, DMACA-HCl or butanol-HCl. The content or composition of each polyphenol was expressed as g/100 mL (if the article describes it). 40 polyphenols were found and build a csv file. Then added to the list of Phenol-Explorer obtaining a total of 144 polyphenols in red wine.

4.1.3 Variables of the database

The final list of 144 compounds was manually curated and cleaned to identify the Pubchem ID, Chebi ID, CTD name, STITCH name, and CAS- number. The SMILES, InchKey, synonims, molecular weight and PubChem ID were retrieved from PubChem database. (https://pubchem.ncbi.nlm.nih.gov/) (Kim et al., 2022).

4.2 Polyphenols diseases, genes, pathways and gene ontology

The network of polyphenols based on different relationships and interactions was built based on the Comparative Toxicogenomics (CTD) database (http://ctdbase.org/) (Davis et al., 2022b). Searching was done by a R package named CTDquerier of Bioconductor (Hernandez-Ferrer & Gonzalez, 2018), using a maximum distance of 1, to identify the exact name as the input .

- Polyphenol-disease: The list of diseases was obtained through the analysis using index_name = “diseases”, including the inferred and the curated associations. The analysis of new therapeutic targets was performed by filtering for disease relations curated by direct evidence as therapeutic/marker/mechanism and with inference score above 0.
- Polyphenol-genes: index_name = “gene interactions”. The analysis of new therapeutic target was done by selecting genes related to Homo sapiens.
- Keggpathways: index_name = “keg pathways”. Bonferroni corrected p-value less than 0.01 was used.
- Gene ontologies: index_name = “go terms”. Bonferroni corrected p-value less than 0.01 was used.

4.3 Protein-Protein interactions

Protein-protein interactions (PPIs) were identified by STITCH database. Version 5.0 (http://stitch.embl.de/) (Szklarczyk et al., 2016), filtering by organism *Homo Sapiens* and confidence score of 0.900, with experimental evidence above 0. The network was built by selecting 20 interactors.

4.4 Screening for target genes in lung cancer

Information on the genes related to lung cancer was acquired from CTD database using the MeSH ID: D008175, using the filter to select the ones with direct evidence data. A total of 87 genes were obtained.

4.5 Common and key targets of polyphenols in red wine and lung cancer

A ggVennDiagram with ggplot R package was run. The selection was obtained from the list of 287 genes by CTD and the list of genes of the red wine polyphenols (9938 genes). The entries found in both lists were used to searched for protein-protein interactions with STRING (http://string-db.org/) version 11.5. (von Mering et al., 2005). The search was limited to “Homo sapiens” and the interactions with probabilistic association confidence score >0.900 were selected.

To explore the molecular mechanism of the red wine polyphenols for Lung cancer, the online functional annotation and enrichment tool DAVID (https://david.ncifcrf.gov/) was used for pathway enrichment, with Kyoto Encyclopedia of Genes and Genomes (KEGG). The organism was set to “*Homo sapiens*” and the results obtained with p-value <0.05, corrected with Benjamin. Finally, the PPI set from STRING was exported as simple textual data format (.tsv), and the TSV file was imported into Cytoscape, version 3.9.1 (https://cytoscape.org/) (Otasek et al., 2019).

For the identification of hub genes it was use the cytoHubba plug-in of cytoscape were it was selected top 100 genes based on the degree and centrality.

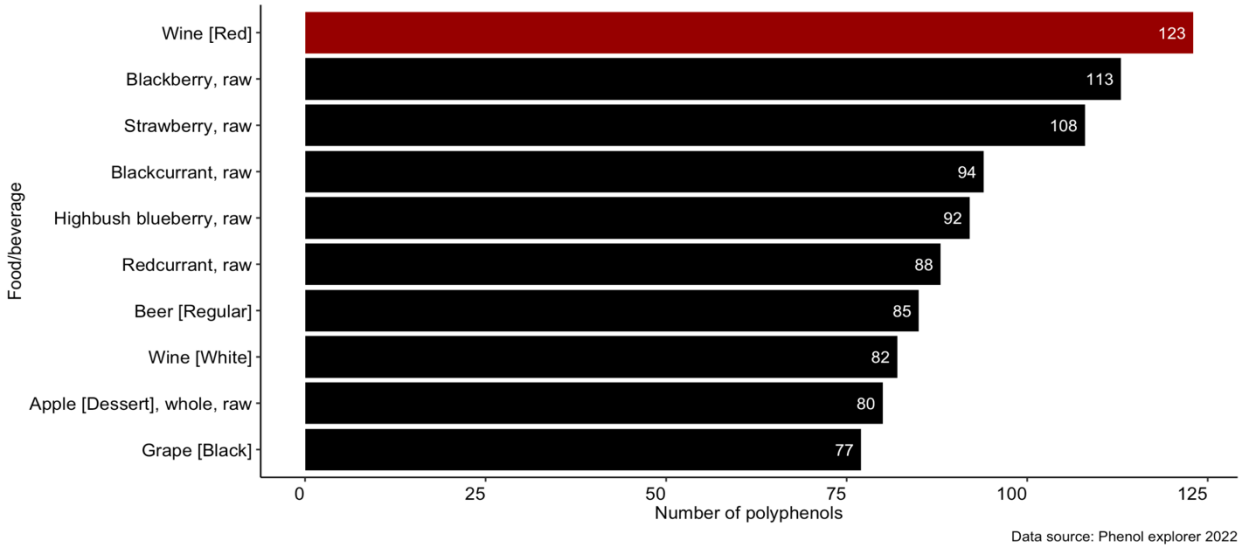
For further details of the methodology, a step by step guide can be found online: https://model3dbio.csic.es/poli_database_redwine

5. RESULTS

5.1 Polyphenols in red wine

The search for polyphenols returned a total of 501 compounds (Table 1. Appendix 8.1) in 458 different food or beverage sources (Table 2. Appendix 8.1) from Phenol-Explorer. After the exploration of the data of phenolic composition in every food/beverage in the database, it was found that the food source with the largest number of different polyphenols is red wine with 123 different polyphenols, followed by blackberry with 113, and strawberry with 108. When compared with other alcoholic beverages, the next one is beer with 83 polyphenols. (Figure 2).

Figure 2: Top 10 Food/beverage with more polyphenol compounds

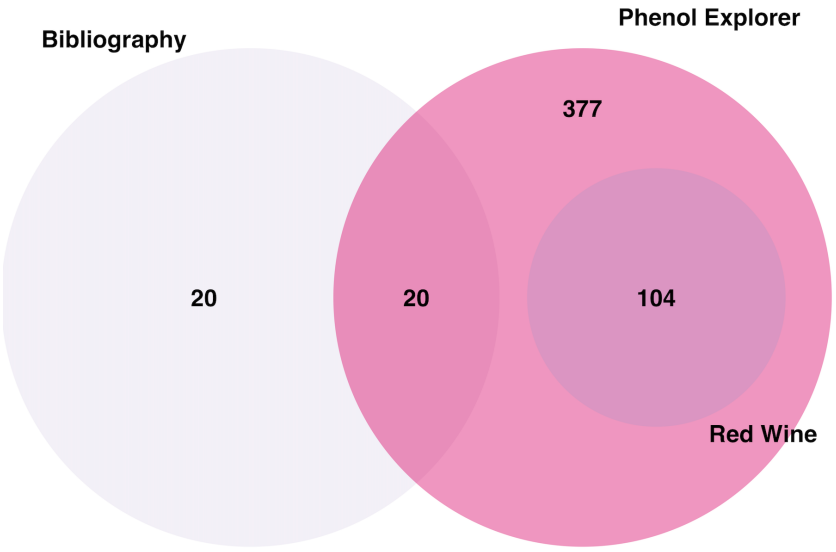


These 123 polyphenols were filtered to get the database of phenolic compounds in red wine (Table 3. Appendix 8.1). After eliminating duplicates and polyphenols that were not representing a unique name (see methodology sec. 4.1.1), a list of 104 polyphenols from Phenol-Explorer (Table 4. Appendix 8.1) was obtained.

Besides using the Phenol-Explorer as data source, a literature search found 40 additional polyphenols. (Table 5. Appendix 8.1). After merging both sources of red wine polyphenols, a final table of 144 polyphenols was obtained, and this is the current number of red wine polyphenols in our database. (Table 6. Appendix 8.1)

Figure 3 shows a Venn Diagram with the intersection of the two abovementioned data sets, where 20 of the 40 polyphenols from the bibliography are actually included in Phenol-Explorer. However, for some reason they are not reported as components of red wine according to the curators. The other 20 polyphenols from the literature are not included in Phenol-Explorer at all.

Figure 3: Source of Polyphenols in Red Wine

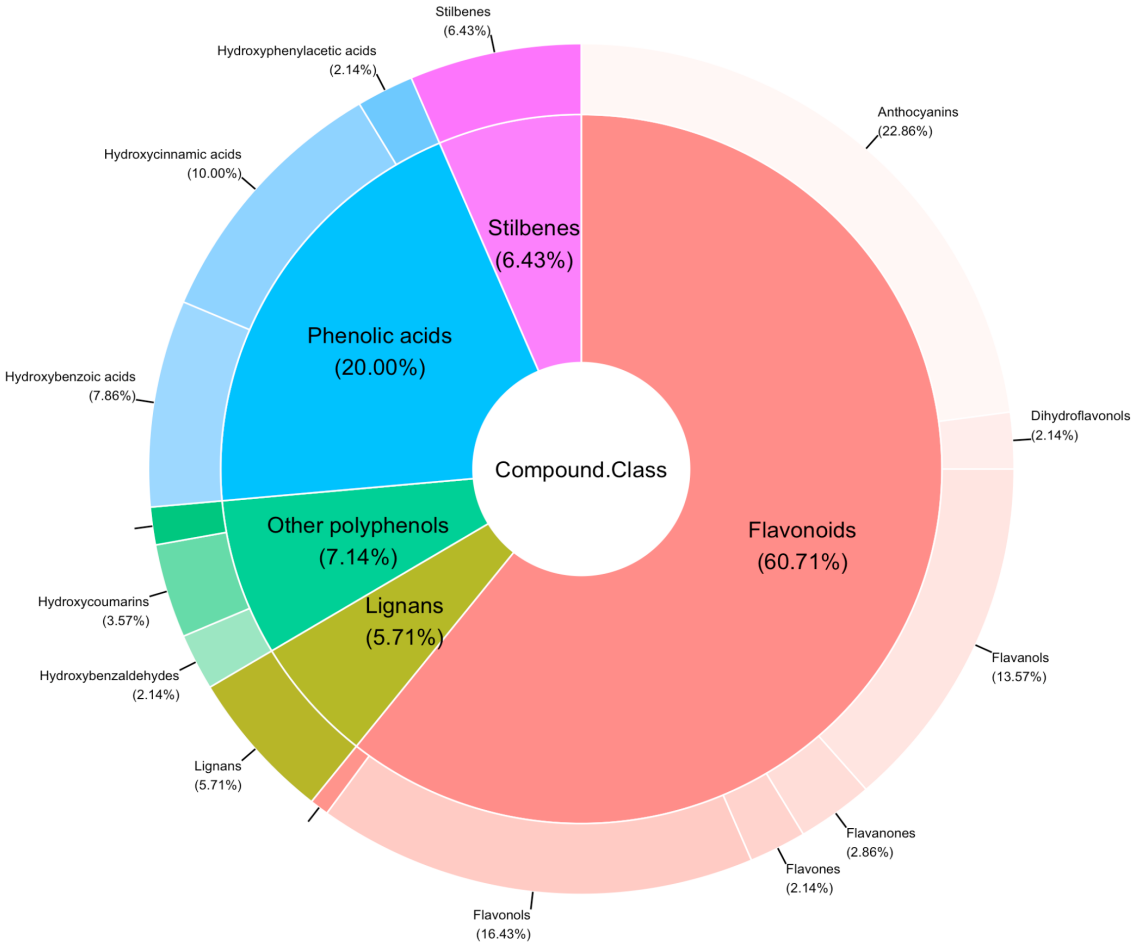


These 144 polyphenols were manually curated to obtain a Pubchem ID, SMILES and InchKey. From these, 140 compounds (Table 7. Appendix 8.1) were identified, but four of them did not match ~~with~~ any Pubchem ID. Having missing IDs makes further analysis difficult, that is why they are going to be listed in the red wine polyphenol database, but they are not going to be used for further analysis.

5.2 Classes and sub-classes of polyphenols in red wine

Figure 4 shows the total number of 140 polyphenols in red wine (Table 7. Appendix 8.1) classified in classes and sub-classes of polyphenols. It can be observed that flavonoids form the most abundant class (60.71% of the total). In second place we have phenolic acids (20.00%), and in third place, other polyphenols (6.94%), followed by stilbenes with 6.25% and lignans with 5.56%. When analyzed by sub-classes, anthocyanins, flavonols and flavanols comprises more than 50% of the total phenolic compounds. Anthocyanins, which represent the 22.86%, are the largest subgroup of wine polyphenols.

Figure 4: Classes & Sub-classes of Red wine Polyphenols



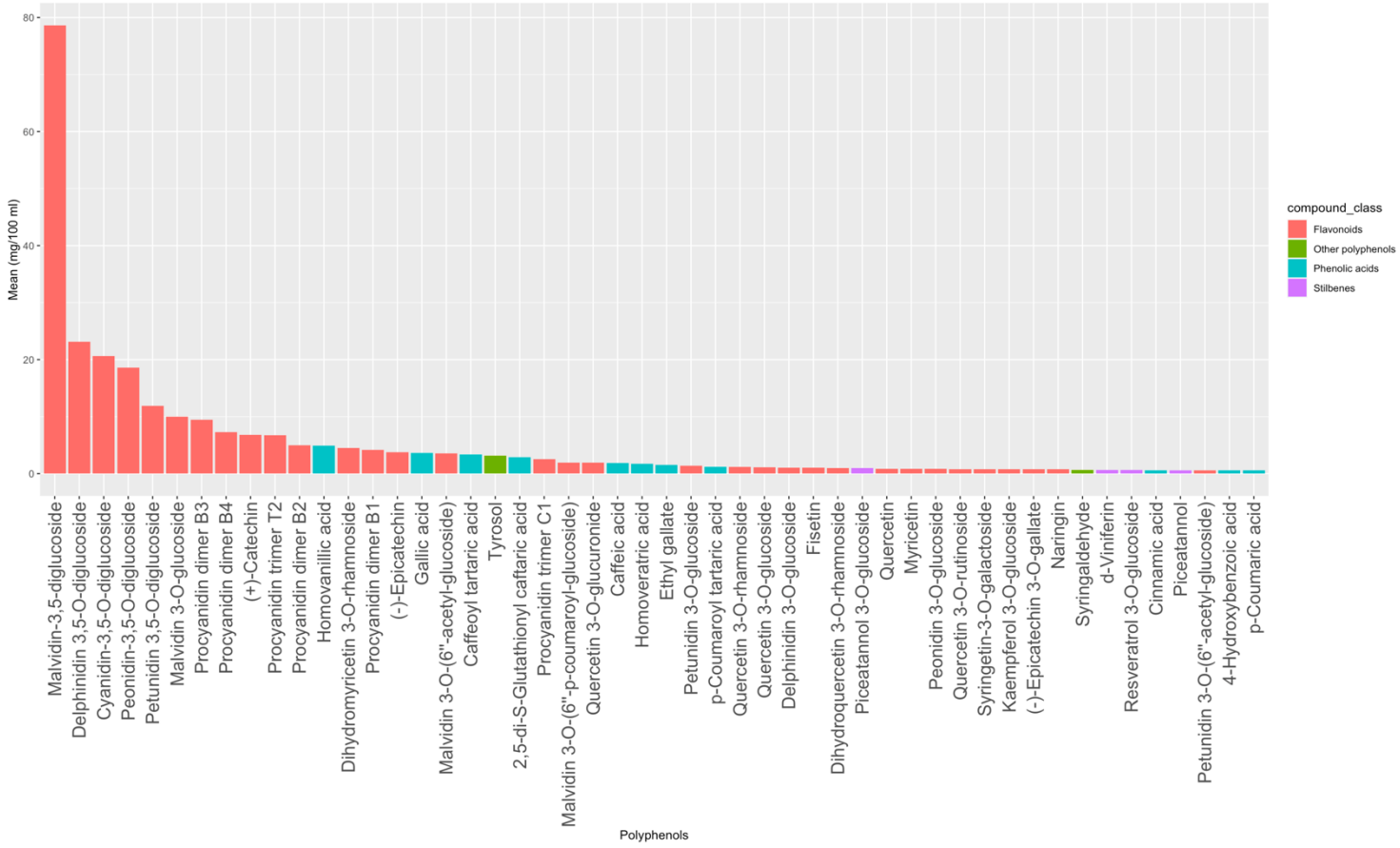
5.3 Content of Polyphenols in red wine by classes

Besides the classification between groups and subgroups, 124 polyphenols out of 140 polyphenols presented information about its concentration in mg/100 mL detected by Liquid Chromatograpy. (Table 7. Appendix 8.1), where it was found a concentration rank between 0.00-78.62 mg/100 mL.

It can be seen in Figure 5, that flavonoids are not only the polyphenol group formed by the largest number of components, but they contain also the polyphenols found at higher concentration in red wine. Malvidin 3,5-diglucoside is the polyphenol with higher concentration, with an average of 78.62 mg/100 mL. Second, delphinidin 3, 5-O-diglucoside with 23.16 mg/100 mL, followed by cyanidin 3,5-O-diglucoside with 20.64 mg/100 mL.

In the group of phenolic acids, the homovanillic acid is the one found at the highest concentration, with 4.92 mg/100 mL. In the class of other polyphenols, tyrosol is the one found at the highest concentration, with 3.12 mg/100 mL, and from stilbenes, piceatannol 3-O-glucoside with 0.94 mg/100 mL.

Figure 5: Top 50 Content of Polyphenols in Red wine by classes



5.4 Genes, diseases, pathways and gene ontologies related to polyphenols

The Comparative Toxicogenomics (CTD) database was used to find the genes, disease, pathways and gene ontologies related to the 140 polyphenols (Table 8 Appendix 8.1), with the R package of Bioconductor “CTDquerier” (Hernandez-Ferrer & Gonzalez, 2018). We found information for 102 compounds of the database (Table 9, Appendix 8.1) and 38 polyphenols were not found in the CDT database (Table 10, Appendix 8.1).

The associations found between the phenolic compounds and diseases, genes, pathway, or gene ontologies can be curated or inferred from interactions (see next section). For example, ~~for~~ the

polyphenol caffeic acid methyl ester is found in CTD but there are not associations that have been curated for this chemical yet.
For the identification of potential therapeutic targets, we have used the ones that have associations with direct evidence, as below described.

5.4.1 Diseases related to polyphenols

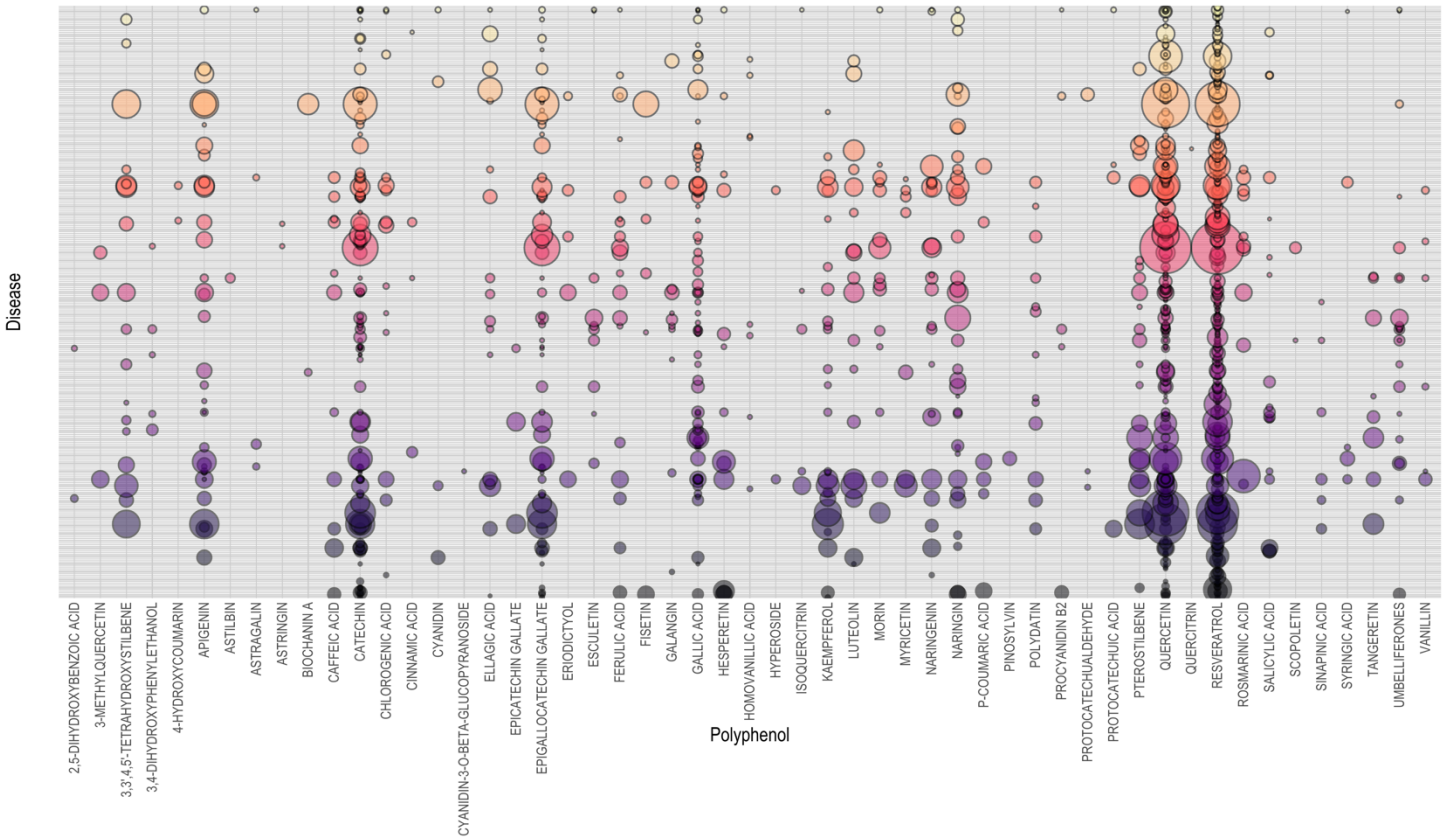
A total of 69039 diseases were related to red wine polyphenols (Table 11, Appendix 8.1) as part of the database created, which includes inferred and curated disease-chemical associations. Curated associations have direct evidence from published literature, while inferred associations are established via curated chemical–gene interactions or disease-gene interactions (indirect associations) (Hernandez-Ferrer & Gonzalez, 2018). Nevertheless, to narrow down the search of new targets, we used a filter where associations that were not supported by direct evidence based on therapeutic, marker or mechanism, were excluded, resulting in 1344 diseases. Then, the inference score was defined as above 0, ending up with 1184 diseases related to 54 polyphenols (Table 12, Appendix 8.1).

Figure 6 shows the red wine polyphenols related with disease, where the size of the dot is proportional to the inference score, ranging from 1.87 to 433.72. The inference score represents the degree of similarity between chemical–gene–disease networks and a similar scale-free random network. The bigger the size of the dot, the higher the score, and the more likely that inference network has atypical connectivity (King et al., 2012). The color of the dots represents a different disease.

The disease with the highest inference score (433.72) is liver cirrhosis, followed by prostatic, breast, colon, lung and stomach neoplasms (see Table 12, Appendix 8.1 for the name of each specific disease, or select the region with the mouse in the Figure 6 of the online database https://model3dbio.csic.es/poli_database_redwine).

Resveratrol, quercetin, epicagallocatechin gallate, naringin and gallic acid are the polyphenols with more dots, bigger sizes, and more different colors, meaning that are the ones related to the largest number and most different diseases.

Figure 6: Polyphenols and Disease



Furthermore, Figure 7 displays the top 25 polyphenols related to the largest number of diseases, where resveratrol is the polyphenol related with more diseases with 285 associations, followed by quercetin with 190 relations, catechin with 103, gallic acid with 73, and epicatechin gallate with 61. This means that a polyphenol can be associate to many diseases. The ones that are associated with only one disease are astilbin, cyanidin-3-beta-glucopyranoside, pinosylvin, and quercitrin (Table 12, Appendix 8.1).

Figure 7: Top 25 Polyphenols related to diseases

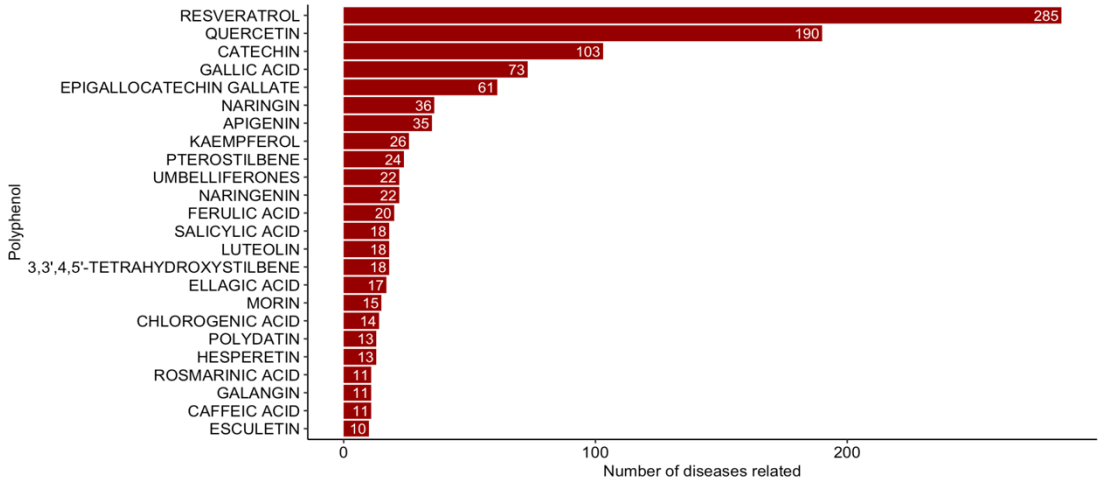
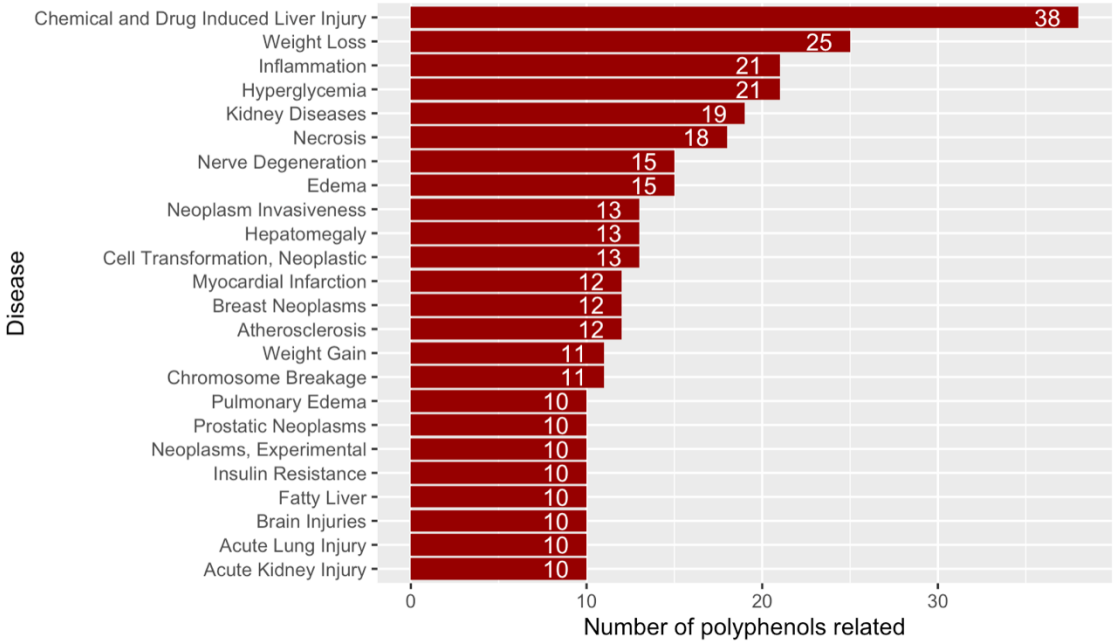


Figure 8 shows the top 25 diseases that are associated with the largest number of polyphenols. It can be seen that “Chemical and drug induce liver injury” is related to 38 polyphenols, “weight loss” to 25 polyphenols, “inflammation” and “hyperglycemia” are associated with 21 each, and “kidney diseases” with 19.

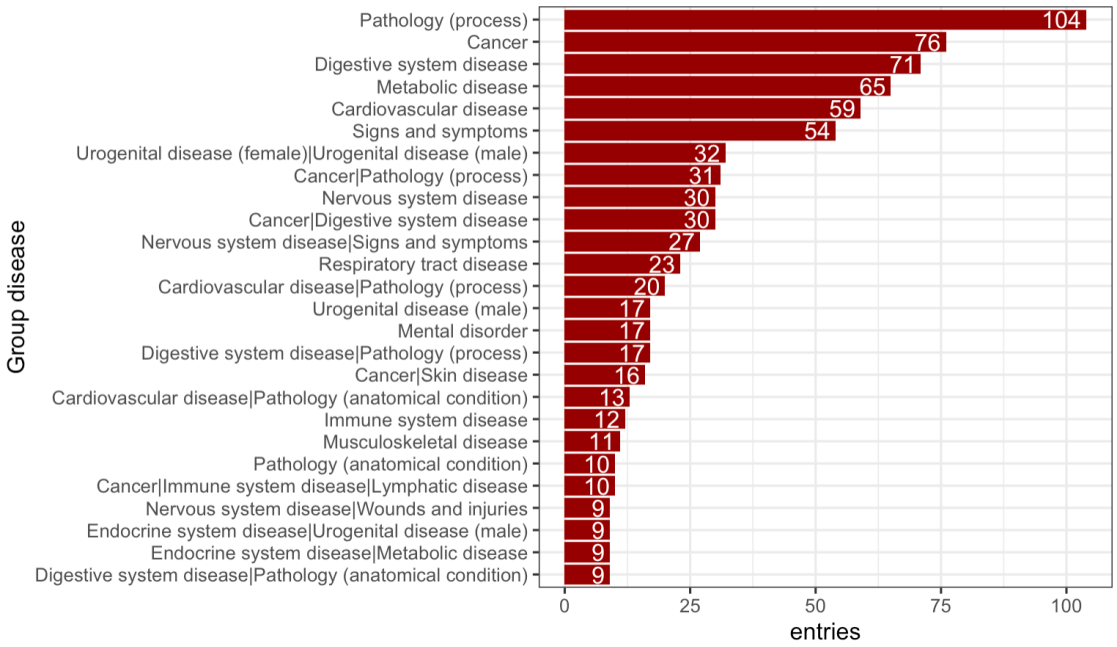
Figure 8: Top 25 Diseases related to Red wine polyphenols



In CTD database it is possible to select group of disease, hence this classification was explored obtaining 129 group disease represented in Figure 9 (Table 13, Appendix 8.1). “Pathology process”, “cancer” digestive, metabolic, and cardiovascular disease are the ones with more entries of disease polyphenols associations.

In the figure are names of groups that sound similar for example “Pathology (process)”/ “Cancer|Pathology(process)” / “Cardiovascular disease|Pathology (process)” where referrers to different groups and hierarchies. For example, in the case of “Pathology process” it refers to the abnormal mechanisms and forms involved in the dysfunctions of tissues and organs, like nerve degeneration and pathologic neovascularization. In “Cancer|Pathology (process)” -that is in the eight position of the rank- is a sub-group of pathological process, that is called neoplastic process where are found the pathological mechanisms and forms taken by tissue during degeneration into a neoplasm, as neoplasm invasiveness, and “Cardiovascular disease|Pathology (process)” is a part of the subgroup of necrosis/isquemia (Davis et al., 2022a).

Figure 9: Top 25 Disease group more related t



5.4.2 Genes related

It was obtained from CTDquerier 45633 chemical-gene curated interactions (table 14, Appendix 8.1), this number includes interactions with different organisms. For this study, it was selected associations related to *Homo sapiens*, ending up with 26,674 interactions, 9939 unique genes and 65 different polyphenols with genes related (Table 15, Appendix 8.1).

In Figure 10, resveratrol, quercetin, catechin epicatechin gallate and gallic acid are the ones with more gene's interactions, as they have more tiles. The darkness of the color expresses the number of times the polyphenol is related to the gene. For example: the gene TNF is related to resveratrol, 223 times, with quercetin 78 times with Luteolin with 27 times. Each time with a different interaction (increase, decrease or affects proteins) (Table 15, Appendix 8.1).

To compare the number of interactions in Figure 11, it is represented the top 5 polyphenols with more interactions. Resveratrol is by far the one with more than 10,892 interactions, followed by quercetin with 6053 interactions, epicatechin gallate with 2743, catechin with 1908 and gallic acid with 722. This is expected since are the polyphenols with more diseases related too.

Figure 10: Polyphenols-genes interaction

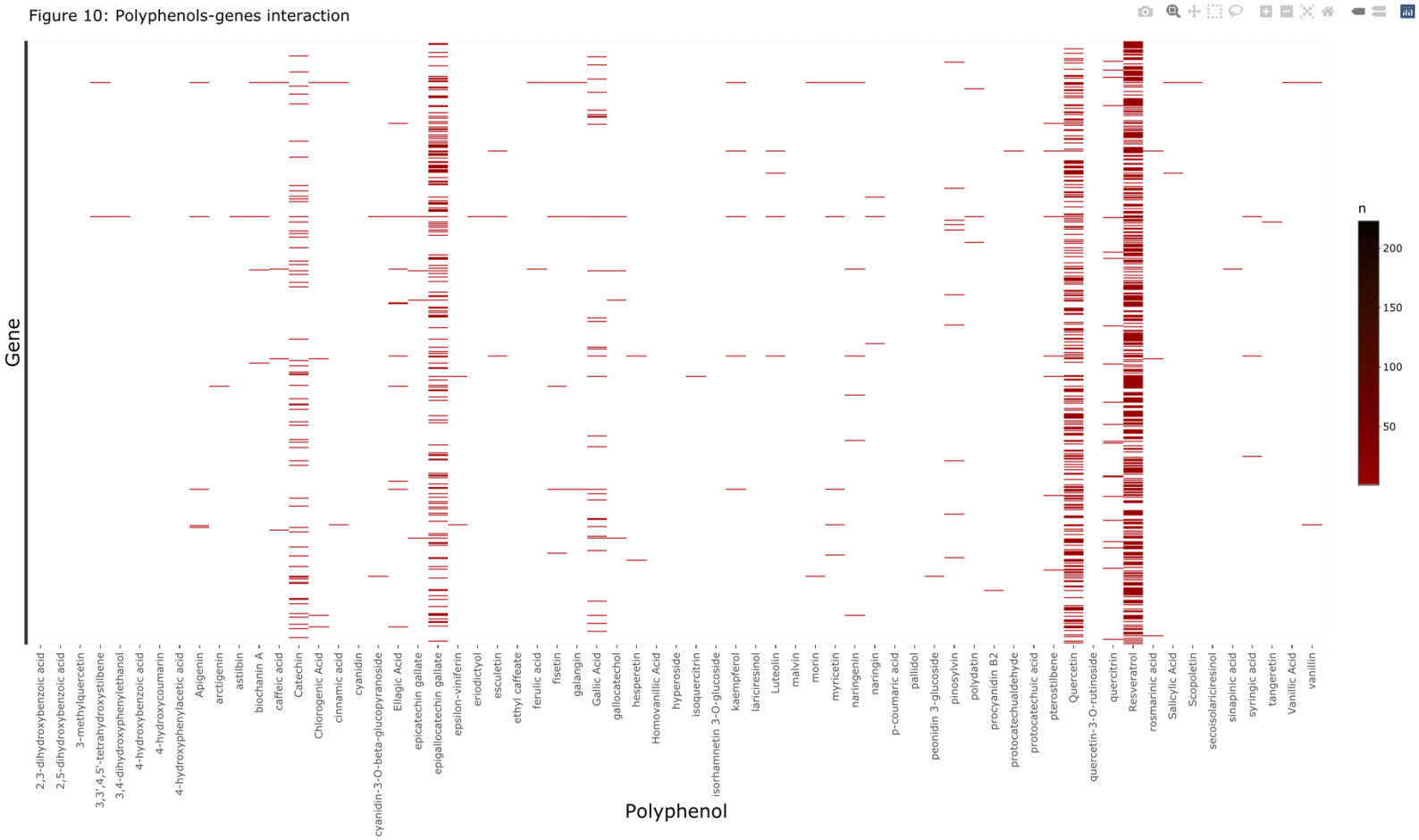
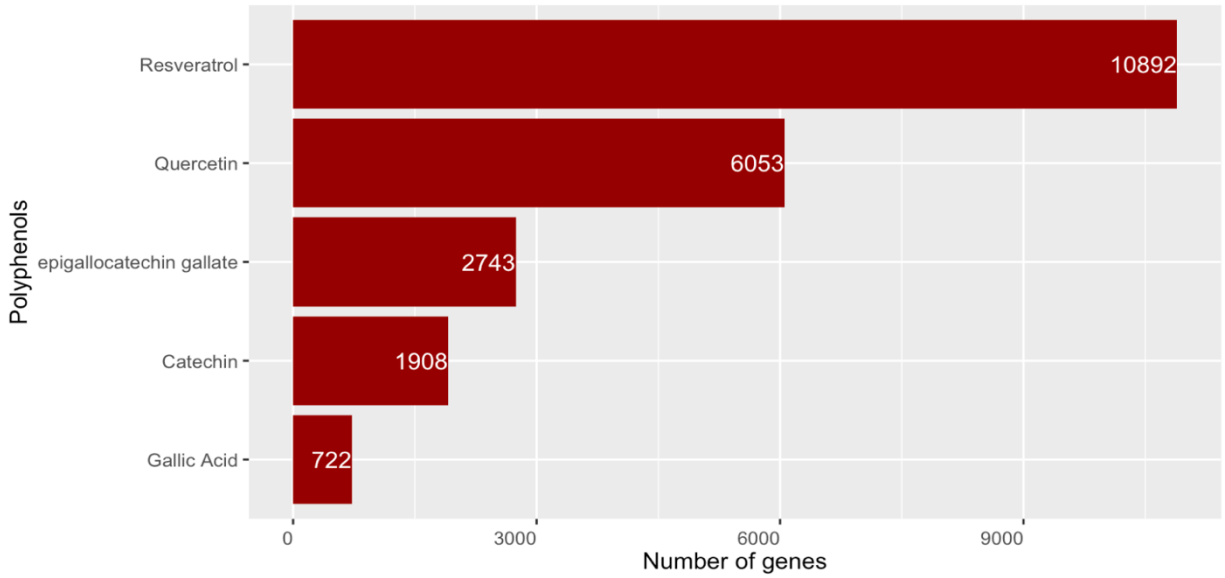
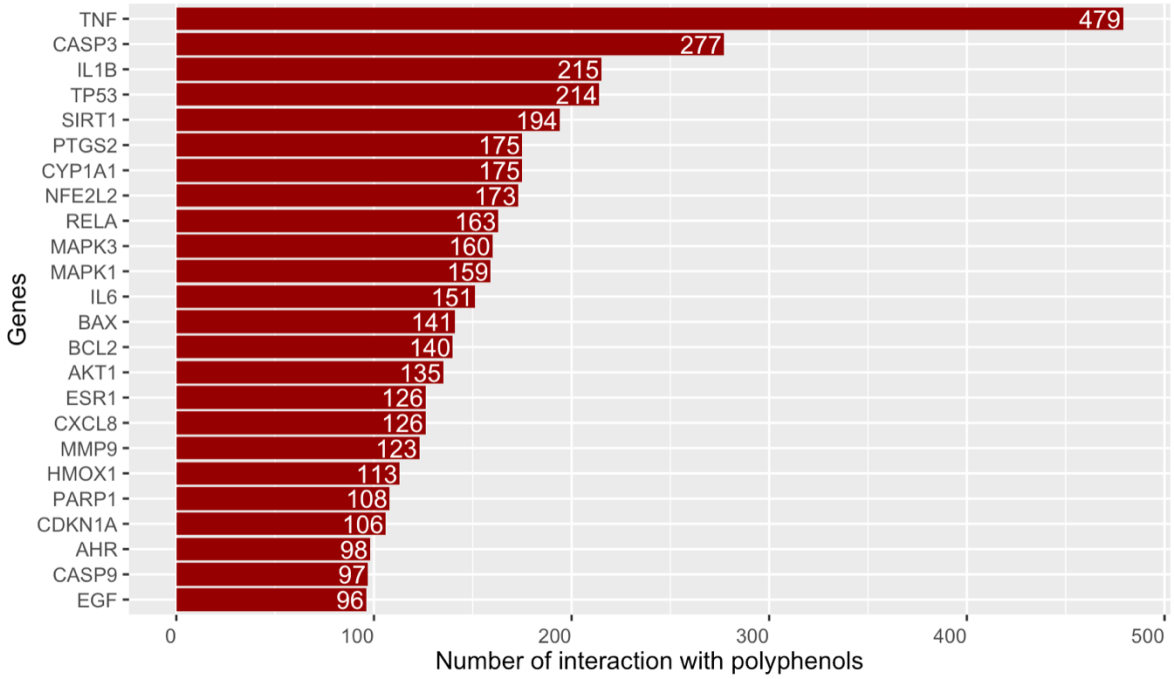


Figure 11: Top 5 Polyphenols with more interactions



Beside the polyphenol with more gene interaction, we also can rank the genes that have more association with polyphenols. In Figure 12, TNF interacts 479 times with 26 different polyphenols producing different interactions, CASP3, 277 times with 34 polyphenols and IL1B 215 times with 18 polyphenols. There are genes that interact with only one polyphenol, for example: A1BG that interacts only with quercetin. (Table 15, Appendix 8.1).

Figure 12: Top 25 Gene interaction with polyphenols



The TNF is the tumor necrosis factor, and has a role in normal physiology, acute inflammation, chronic inflammation, autoimmune disease, and cancer-related inflammation (Chu, 2013). CASP3, which is the caspase 3, is a lysosomal enzyme involved in the apoptotic pathway. (Kashyap et al., 2021) and IL1B the Interleukin 1 beta, this cytokine is an important mediator of the inflammatory response, and is involved in a variety of cellular activities, including cell proliferation, differentiation, and apoptosis (Lopez-Castejon & Brough, 2011). These findings are correlated with the diseases that have been more related with polyphenols in the previous section and the role of the polyphenols as anti-inflammatory that has been studied for several authors. (Baur & Sinclair, 2006; Birrell et al., 2005; Chen et al., 2005; Elmali et al., 2005).

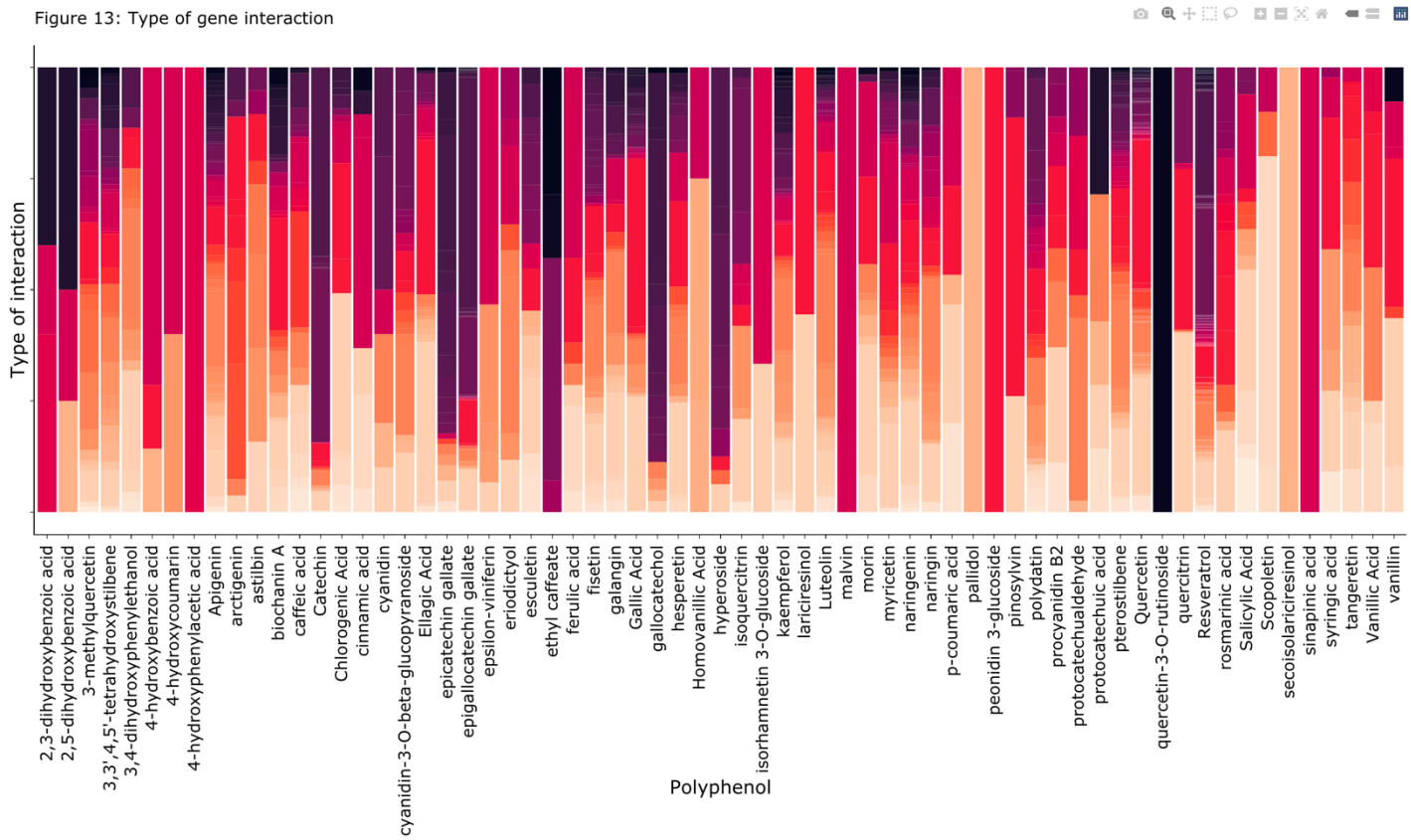
Additionally, it was represented the type of interactions of the genes, in table 15 (Appendix 8.1) can be seen in the “interaction” column, binary associations, involving one chemical with one gene or protein, (for example: “Catechin results in decreased expression of ABCG2 protein”) and

sentences in brackets that represent nested events in more complex interactions (for example: apigenin inhibits the reaction [Brefeldin A results in increased phosphorylation of ERN1 protein]). In the data base were identified 549 different interactions, that characterizes common physical, regulatory, and biochemical interactions between chemicals and genes.

In figure 13 it is possible to see the proportions of them. The different colors, represent, the operators that describe the degree of the chemical effect. The light orange color represents the operator “increase”, the red color “decrease”, and the dark purple the degree is not specified like, “affect-binding” or, “affect cotreatments”. (For a specific interaction see table 15 (Appendix 8.1), or in the web (https://model3dbio.csic.es/poli_database_redwine), with the mouse and select a region in the figure 13).

The stack bars with more different color range are the ones with more interaction like resveratrol, quercetin, and catechin, which is expected, because they have the mayor number of genes associations. And the ones with single color have just one type of interaction, like malvin that the only interaction is -decreases^activity- of two proteins CA1 and CA2 or pallidol that -increases^activity- of the NFE2L2 protein (Table 15 (Appendix 8.1).

Figure 13: Type of gene interaction



5.4.3 Pathway related

CTD database searches pathways from two sources Kegg pathways and Reactome pathways. And were identified 13732 polyphenol-pathways relations with 65 red wine polyphenols (Table 16, Appendix 8.1). All the pathways are enriched with a Bonferroni p-value less than 0.01.

In Figure 14, the smaller p-value (represented with a darker red tile) the stronger is the evidence. Is possible to distinguish that resveratrol, quercetin, epigallocatechin gallate, catechin and gallic acid are the polyphenols with more interactions with Pathways. This is expected due to these polyphenols are the ones with more disease and genes interactions.

In the superior area of the Figure 14 it is possible to see almost a horizontal dark red line, which could be interpreted as there are pathways thar are related with almost with the polyphenols. To identify these pathways were ranked and presented in Figure 15.

Figure 14: Polyphenols and pathways

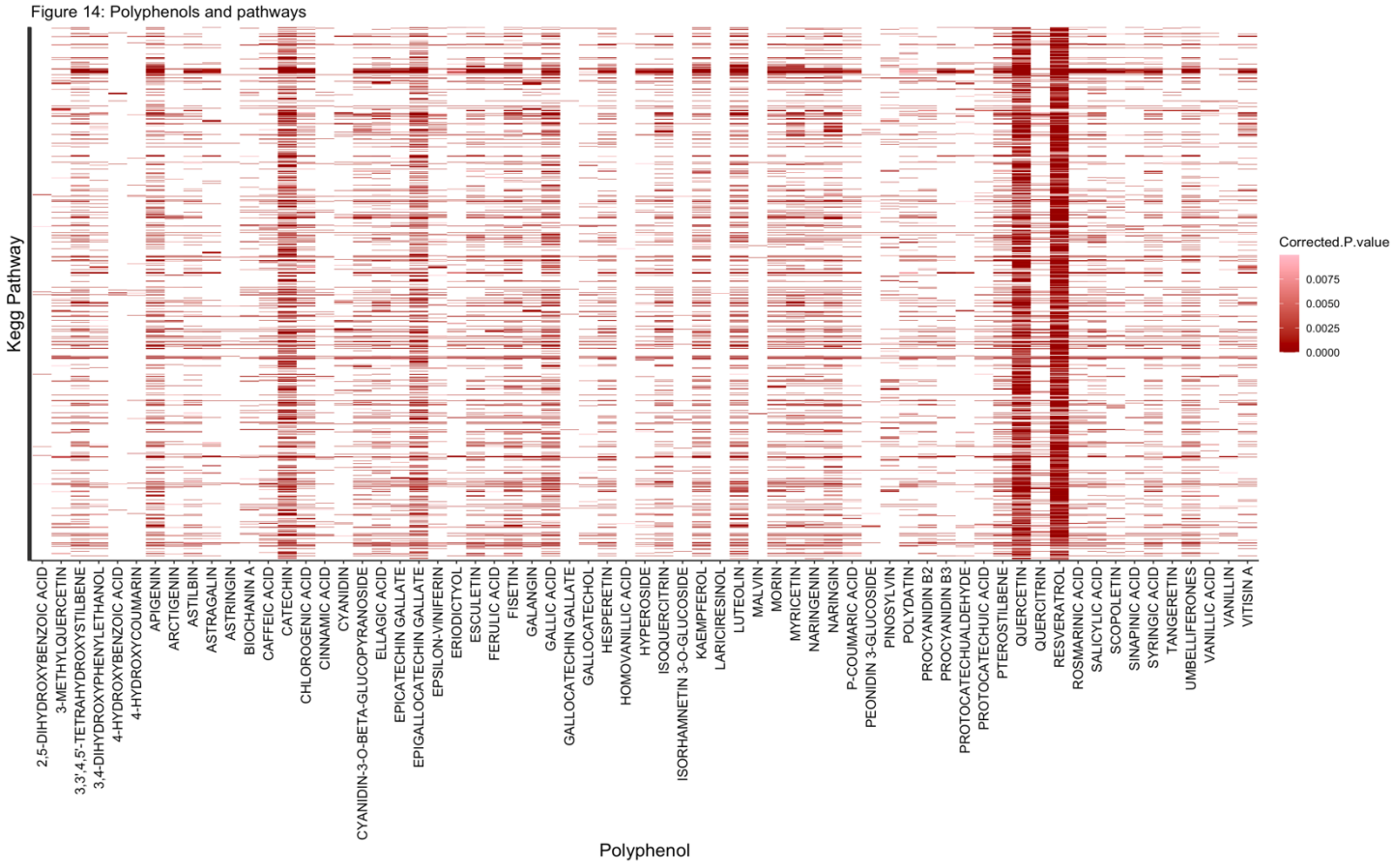
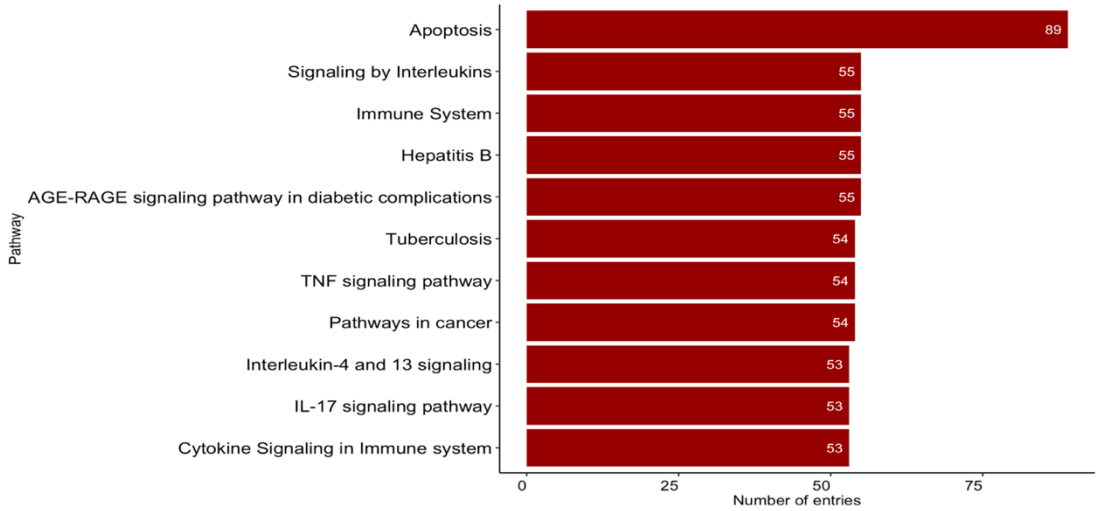


Figure 15 displays that most of the polyphenols are interacting with “Apoptosis” pathway having 89 entries (The number of entries refers to the number of times, there is an interaction between polyphenol and pathways), followed by AGE-RAGE signaling pathway in diabetic complications, Hepatitis B, Immune System and Signaling by interleukins with 55 relations each of them. These pathways are expected since the genes like TNF, CASP3, and IL1B, are involved in inflammation and apoptosis and are the ones that interact more with wine polyphenols.

Figure 15: Top 10 Pathways

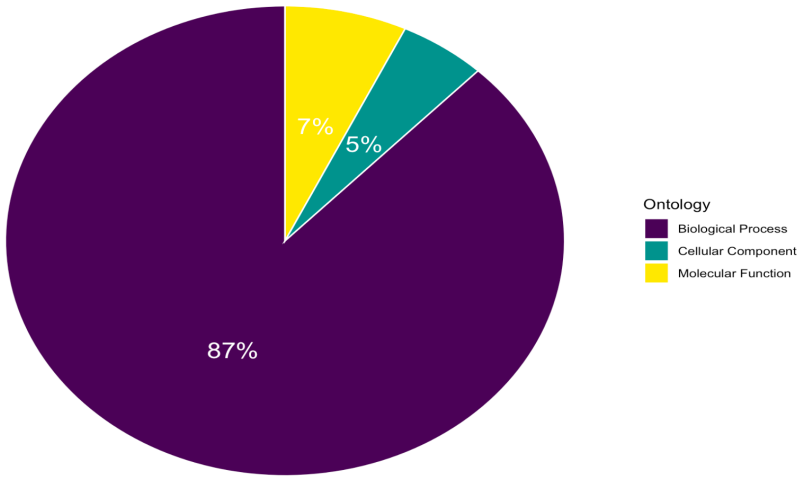


5.4.4 Gene ontologies (GO)

There are 66 polyphenols related with 5482 gene ontologies that generates 64276 entries (Table 17, Appendix 8.1), the curated interactions between chemical and genes, have gene ontology annotations that give information about their associated biological processes, molecular functions, and cellular components. These ontologies are enriched with a Bonferroni p-value less than 0.01.

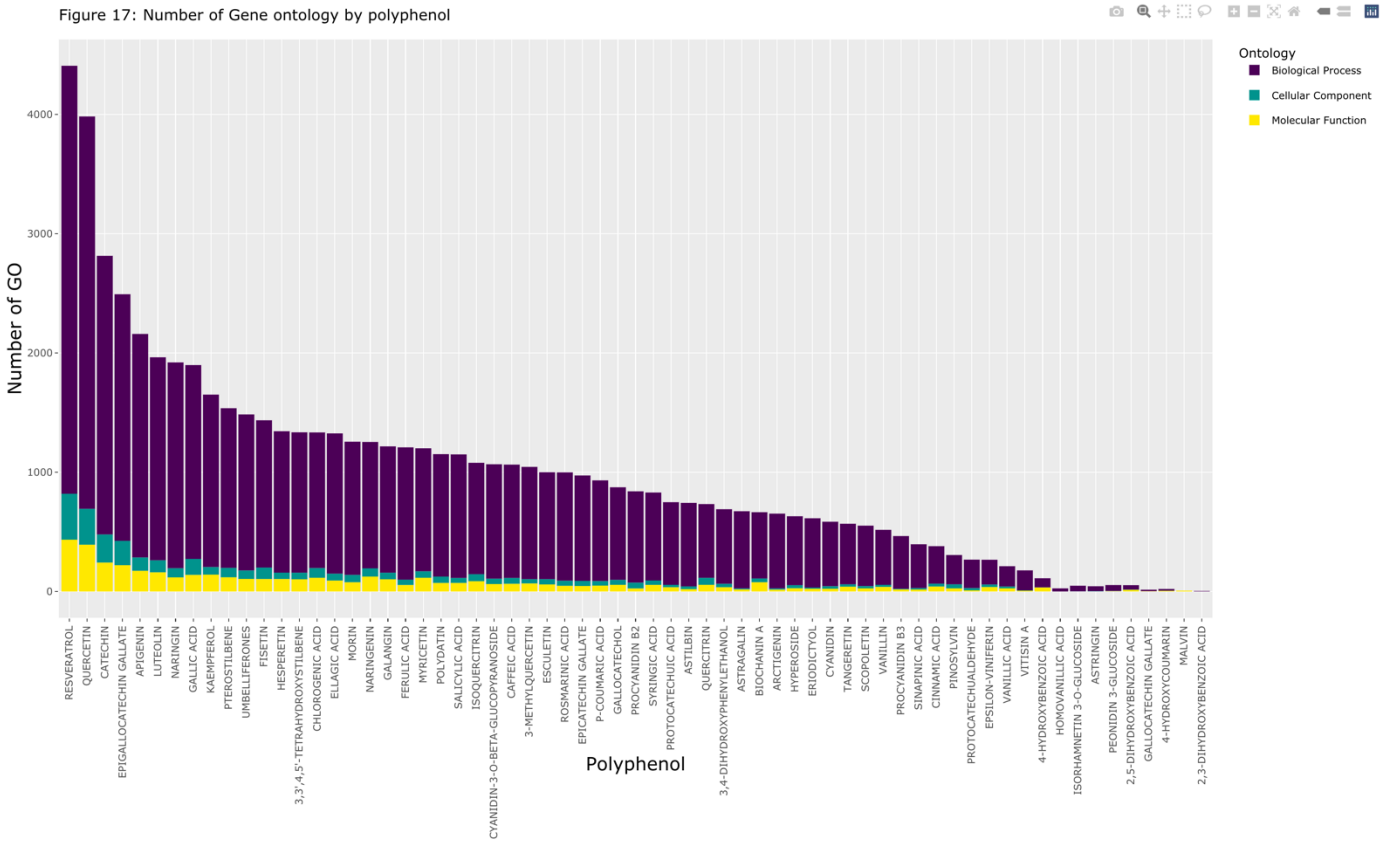
Figure 16 shows that 87% of the ontologies in red wine polyphenols are related to biological process, 7% to Molecular function and 5% to Cellular component. In figure 17, it can be seen a global distribution of the gene ontologies, where it can be identifying the ones with fewer GO association like malvin with seven, or 2,3-hydroxybenzoic acid with only four GO related. And the polyphenols with more gene ontologies, is resveratrol that have 4,409 GO, quercetin with 3984 GO and catechin with 2815 GO. (For a specific GO see table 12, or in the web, https://model3dbio.csic.es/poli_database_redwine , with the mouse and select a region in the figure 17).

Figure 16: Percentage of Ontologies related to RW polyphenols



I

Figure 17: Number of Gene ontology by polyphenol



Likewise, in Figure 18 there is a tile graph where shows the distribution of GO, and as in Figure 17, the polyphenols that stands out are resveratrol, quercetin, epicatechin gallate, catechin and apigenin. We can identify are certain horizontal lines that shows the Gene ontologies that have more associate with the red wine polyphenols. To identify these pathways were ranked and presented in Figure 19 (For a specific GO see table 12, or in the web, https://model3dbio.csic.es/poli_database_redwine , with the mouse and select a region in the figure 18).

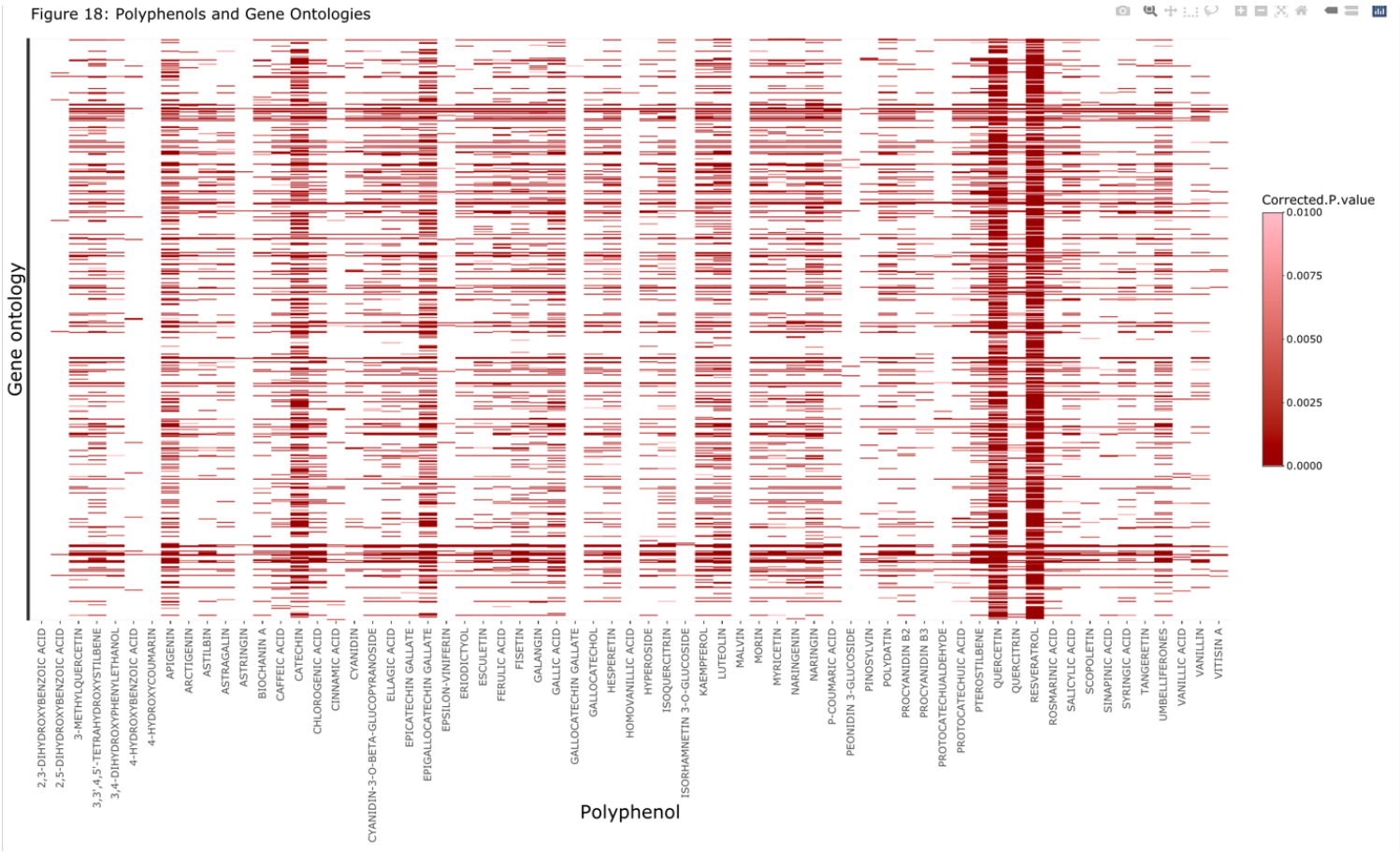
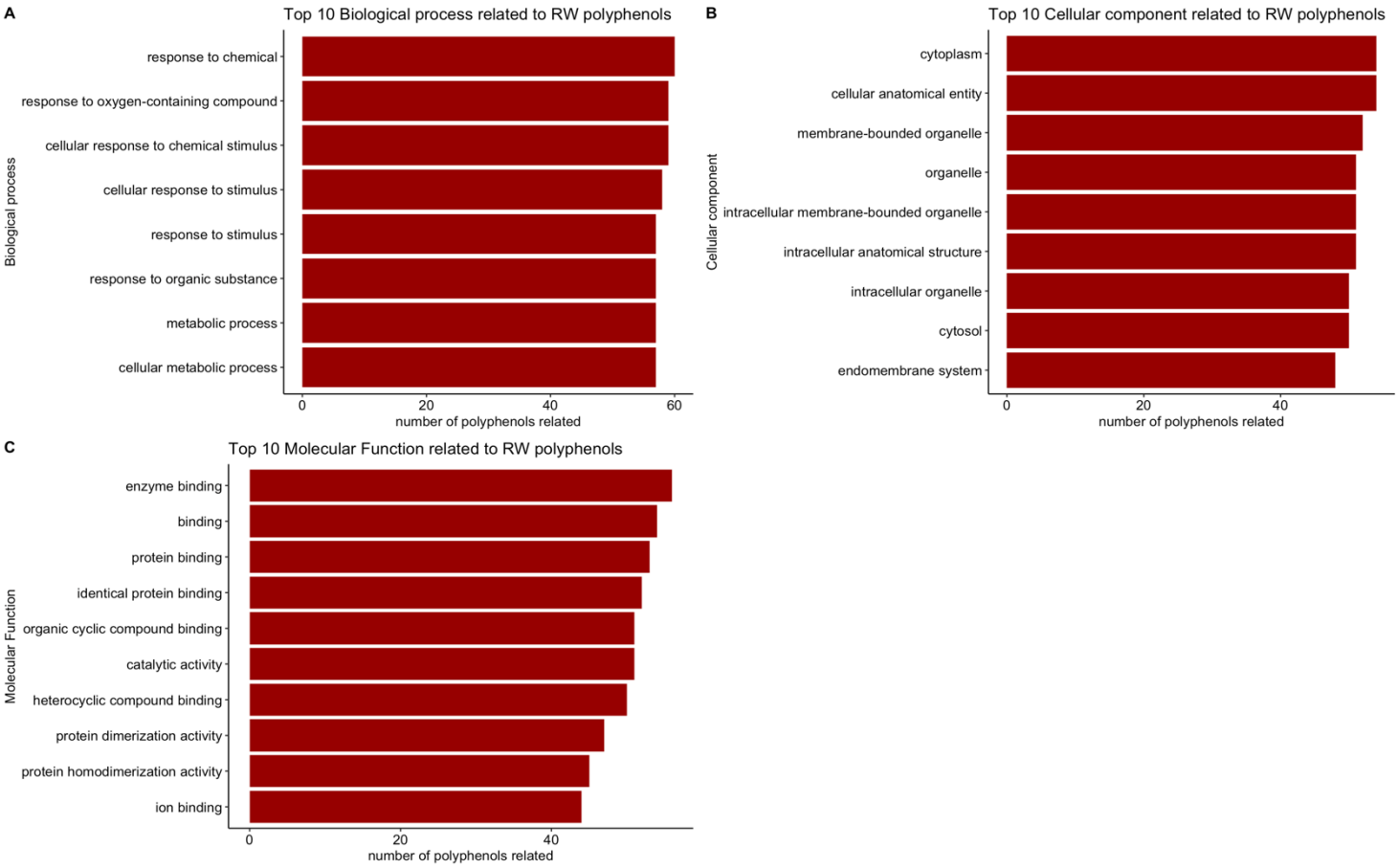


Figure 19 is divided in three parts: were in (A) The biological process, which represents a specific objective that the organism is genetically programmed to achieve, more related to polyphenols are the “response to chemical”, “response to oxygen containing compound” and “cellular response to chemical stimulus”, in case of the Cellular Component (B), that means a alocation, relative to cellular compartments and structures, occupied by a macromolecular machine when it carries out a molecular function. More associate are the “cytoplasm”, “cellular anatomical entity” and “membrane-bounded organelle”. Finally (C) the molecular function, which refers to an action, or

activity, that a gene product (or a complex) performs, more related is “enzyme binding”, “general binding” and “protein binding”.

Figure 19 Gene Ontologies (A. Biological Process, B. Cellular component, C. Molecular Function)



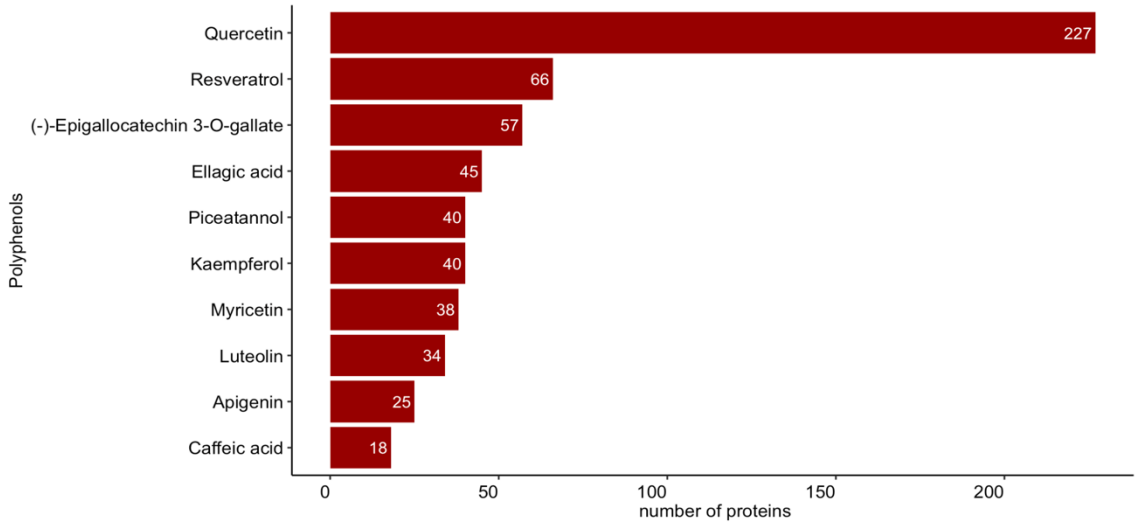
5.5 Protein Binding Information

STITCH is a database that shows interaction chemical protein. For red wine polyphenols were retrieved 14583 interactions between protein and 96 polyphenols, this number includes the inferred and the direct associations (Table 18, Appendix 8.1).

For the study of new targets, it was filtered to use only the ones with direct evidence (experimental 0<) where there are 764 interactions and 42 of the 140 polyphenols, and 466 related proteins (Table 19, Appendix 8.1).

In Figure 20, it is presented the top 10 polyphenols with more interactions with proteins. Where quercetin is in the number on position with 227 proteins association, resveratrol with 66, epicagallocatechin gallate (also named Epigallocatechin 3-O-gallate) with 57, and Ellagic acid with 45 proteins. These is expected, since these are the phenolic compounds with more disease, genes, and pathways associations. There is a group of eleven polyphenols that only one protein target (Table 19, Appendix 8.1).

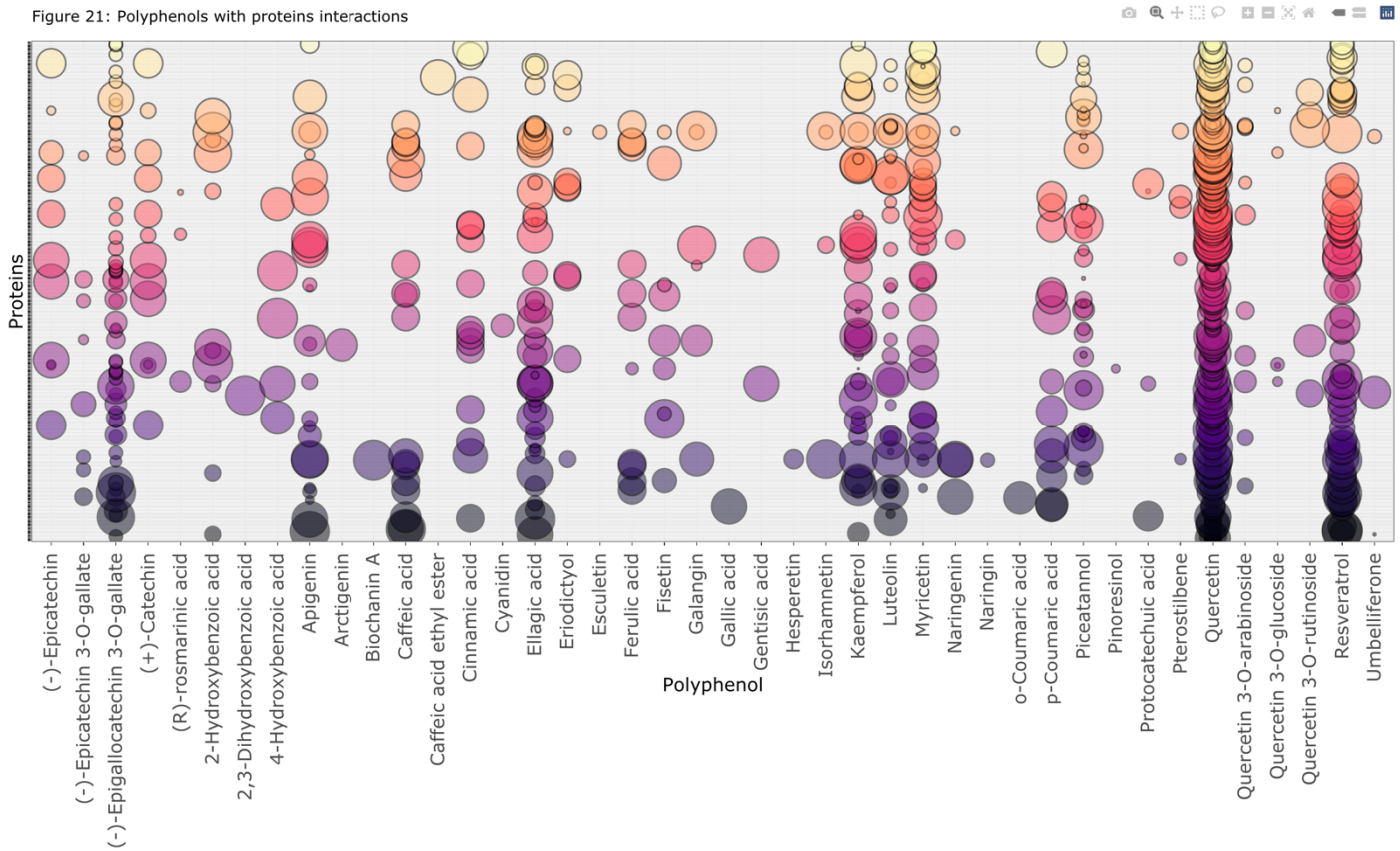
Figure 20: Top 10 Polyphenols and number of proteins interaction



In Figure 21 there is a general exhibition of the interactions between polyphenols and proteins, the size of the dots, represents the combined score. The combined score computed by STITCH is the combination of probabilities from experimental, prediction, database, text mining evidence. (von Mering et al., 2005). The rank of the score is from 151 to 998.

Resveratrol and quercetin shows the greater number of dots and the bigger combined score, and pinoresinol or narigin the ones with fewer number of interactions.

Figure 21: Polyphenols with proteins interactions



An important information that can be analyze using protein binding information is the PPIs, and in Figure 22 is represented the network of the polyphenols based on experimental evidence, with 0.900 confidence and 20 more interactors. These interactors show proteins like Cytochrome enzymes (CYPs) which are isozymes in human liver microsomes that have an important role in the oxidative metabolism of endogenous compounds and clinical drugs. Or AKT1 and AKT kinase, which is one of 3 closely related serine/threonine-protein kinases (AKT1, AKT2 and AKT3) and regulates processes like metabolism, proliferation, cell survival, growth, and angiogenesis (Nicholson & Anderson, 2002; Rönnstrand, 2004). In the network also it is possible to see that there is interaction between compounds like kaempferol with quercetin, and protein-protein interaction for example between cytochromes.

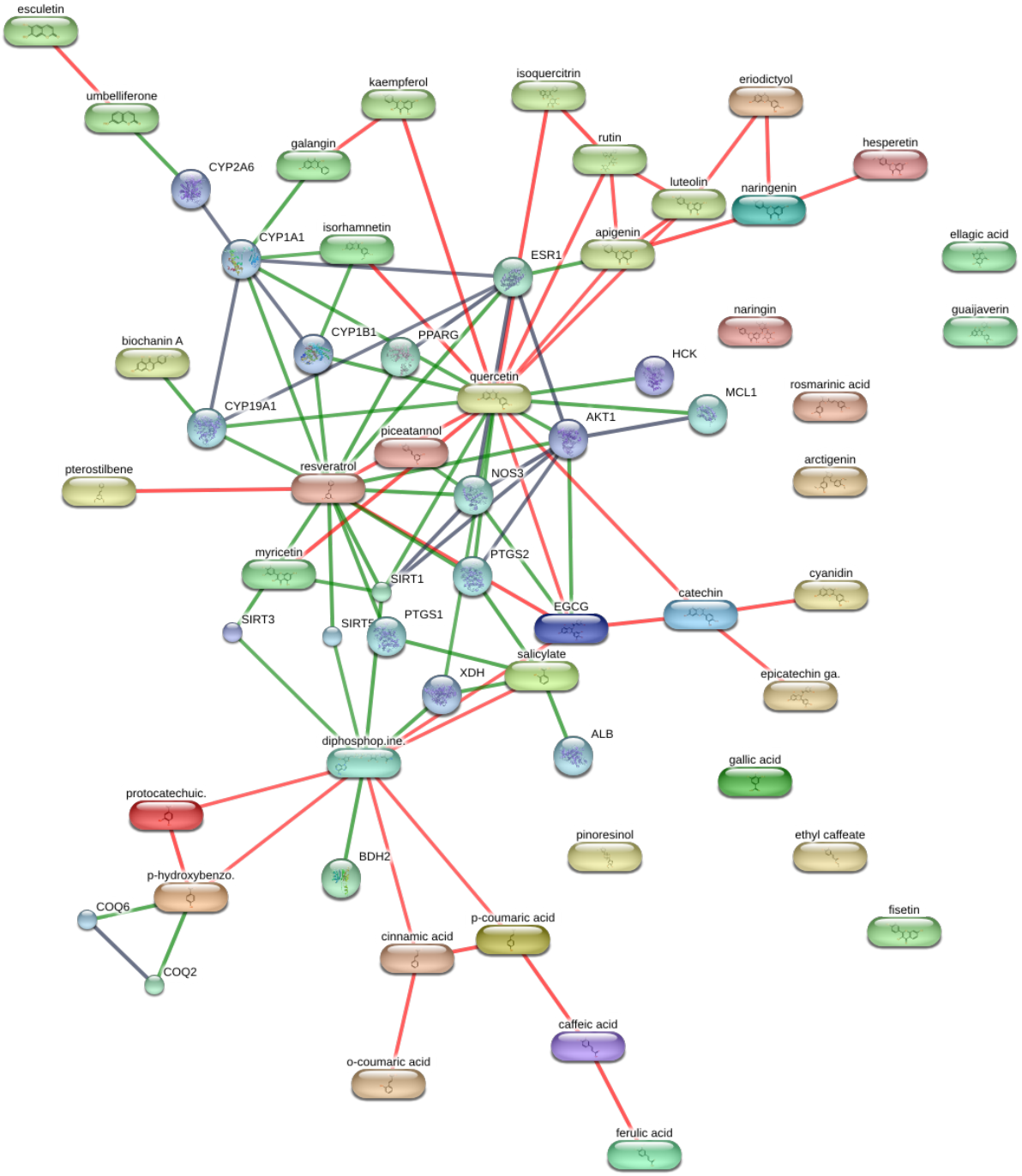


Figure 22: Chemical Protein Interaction The red lines show the interaction between chemical, the grey lines interaction protein-protein, and green lines, chemical protein. The shape of the objects has the following meaning: circular proteins where the size indicates if it is big that the structure is known and if it is small the structure is not known for example COQ6. Rectangular shape is the chemical.

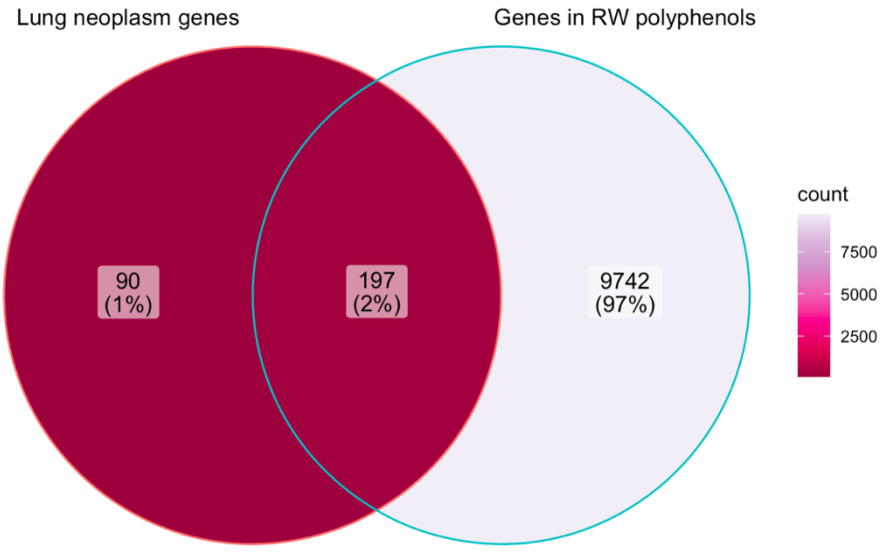
5.6 Example of lung Cancer

To test the database created with red wine polyphenols and find new therapeutic targets it was selected the disease lung neoplasm (MESH:D008175), which is the second leading cause of cancer-related mortality globally, with an overall five-year survival rate of 22.4%. (Lung and Bronchus Cances, n.d.).

For the identification of target, the first step is to identify the genes related to lung neoplasm. Using CTD database were found 287 genes (Table 20, Appendix 8.1), related with direct evidence (therapeutic/marker/mechanism), also it was filter the genes, in the Red Wine database created, related to the disease finding a total of 9742 genes (Table 15, Appendix 8.1).

In the next Venn diagram (Figure 23) is presented the intersection between the genes related to the disease and the genes related to polyphenols in red wine. There are 197 genes shared between genes associated to lung neoplasm and to 50 red wine polyphenols (Table 21, Appendix 8.1).

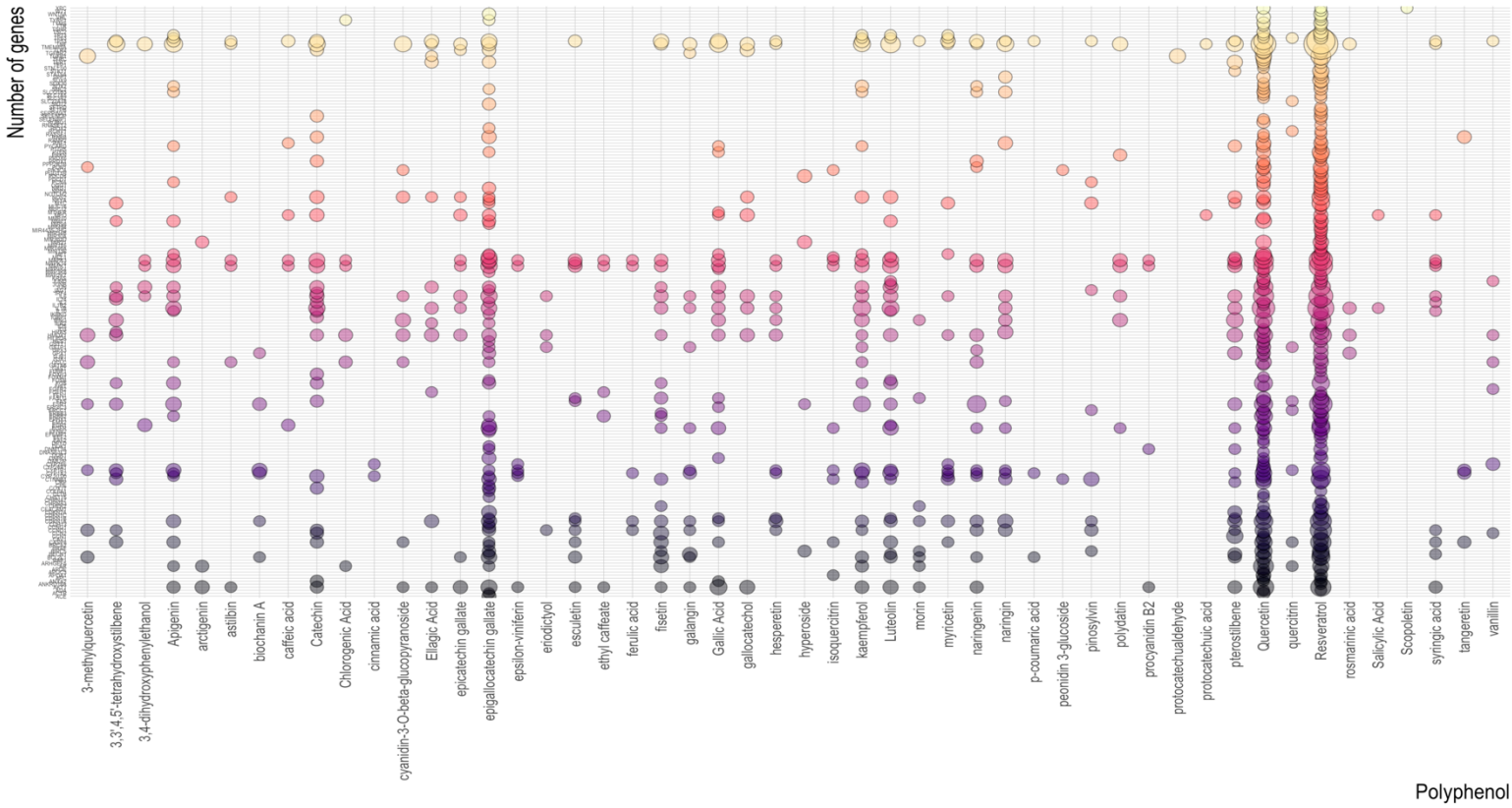
Figure 23: Venn Diagram of Genes related to lung cancer



In Figure 24 is display the list of the name of the 50 polyphenols, and the 197 genes, where the size of the circles represents the count of that gene's interaction with that polyphenol. For example, resveratrol, with TNF gene interacts 223 times. (For a specific interaction see table 21 (Appendix 8.1), or in the web (https://model3dbio.csic.es/poli_database_redwine), with the

mouse and select a region in the figure 24). Also, it is visible that quercetin and epicatechin gallate have high number of lung cancer genes interaction. These findings correlates with studies that have found possible action in lung cancer, for example, quercetin can suppress growth of lung cancer cells as an aurora B inhibitor both in vitro and in vivo (Xingyu et al., 2016). Epicatechin gallate has found that inhibits the stemness and tumourigenicity of human lung cancer cells by inhibiting AXL. A receptor tyrosine kinase that stimulates cell softening and motility in human lung cancer cells. (Namiki et al., 2020)

Figure 24: Red Wine polyphenols related with lung cancer



After the examination of the gene's association, it is important to identify the proteins-protein interaction to identify hub proteins and possible targets. Using the list of 197 genes and STRING database, the resulting network can be seen in the Figure 25. Showing PPI with 0.900 confidence, and no interactors added.

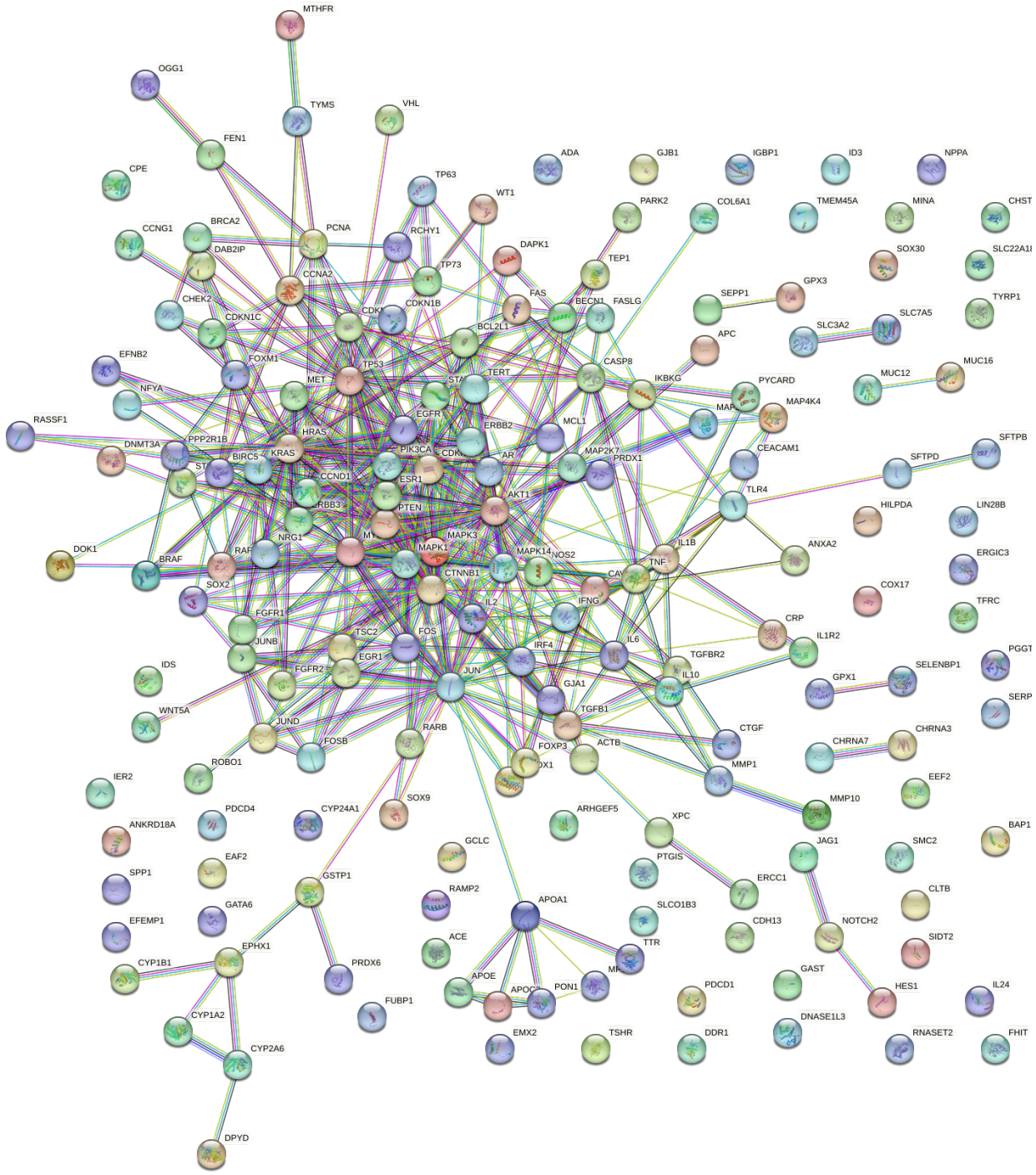


Figure 25. PPI Lung cancer-red wine polyphenols network.

Edges represent protein-protein associations	Known Interactions	Predicted Interactions	Others
associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other.	*from curated databases* *experimentally determined*	*gene neighborhood* *gene fusions* *gene co-occurrence*	*textmining* *co-expression* *protein homology*

It is a network, with known and predicted interactions and in order to identify possible targets it is necessary to find the pathways that are enriched. To perform the enrichment was used DAVID database, finding a list of pathways for 195 of the 197 genes

Figure 26 shows in ascendent order the ones with smaller p-value, (corrected using Benjamin). As expected, there are enrichment for pathways related to cancer for example, proteoglycans in cancer, Breast, prostate, hepatocellular bladder and non-small cell lung cancer (NSCLC) (this is a type of lung neoplasm)

Figure 26: DAVID Enriched pathways (Lung-polyphenol genes)

Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Homo sapiens
195 DAVID IDs
Options

Rerun Using Options | Create Sublist

1293 chart records

Download File

1, 2

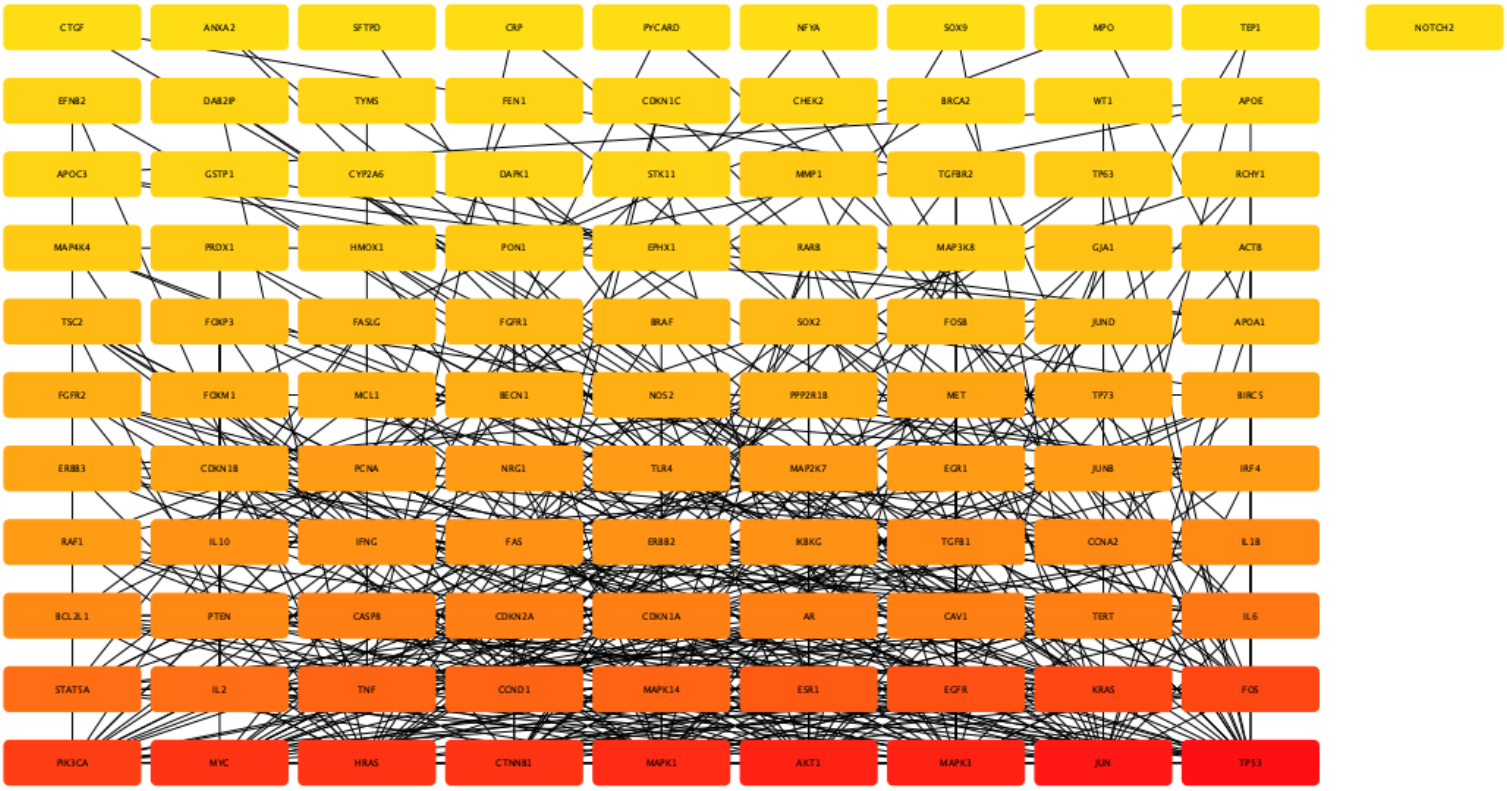
Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	KEGG_PATHWAY	Pathways in cancer	RT		53	27.2	2.1E-24	5.4E-22
	GOTERM_BP_DIRECT	positive regulation of gene expression	RT		38	19.5	2.6E-21	8.2E-18
	GOTERM_BP_DIRECT	negative regulation of apoptotic process	RT		37	19.0	3.3E-20	5.1E-17
	GOTERM_BP_DIRECT	positive regulation of transcription from RNA polymerase II promoter	RT		51	26.2	1.3E-18	1.3E-15
	KEGG_PATHWAY	Proteoglycans in cancer	RT		30	15.4	7.7E-18	8.1E-16
	GOTERM_BP_DIRECT	negative regulation of gene expression	RT		28	14.4	8.1E-18	5.5E-15
	GOTERM_BP_DIRECT	positive regulation of cell proliferation	RT		35	17.9	8.8E-18	5.5E-15
	KEGG_PATHWAY	Chagas disease	RT		23	11.8	9.5E-18	8.1E-16
	GOTERM_BP_DIRECT	regulation of cell cycle	RT		28	14.4	1.5E-17	7.9E-15
	KEGG_PATHWAY	Hepatitis B	RT		27	13.8	1.9E-17	1.0E-15
	KEGG_PATHWAY	Breast cancer	RT		26	13.3	2.0E-17	1.0E-15
	GOTERM_BP_DIRECT	negative regulation of cell proliferation	RT		32	16.4	2.8E-17	1.2E-14
	GOTERM_BP_DIRECT	aging	RT		23	11.8	3.0E-17	1.2E-14
	GOTERM_BP_DIRECT	response to estradiol	RT		19	9.7	4.3E-17	1.5E-14
	KEGG_PATHWAY	Hepatocellular carcinoma	RT		27	13.8	4.8E-17	1.9E-15
	KEGG_PATHWAY	Prostate cancer	RT		22	11.3	5.1E-17	1.9E-15
	GOTERM_BP_DIRECT	response to drug	RT		26	13.3	1.5E-16	4.6E-14
	KEGG_PATHWAY	MicroRNAs in cancer	RT		34	17.4	2.0E-16	6.5E-15
	GOTERM_BP_DIRECT	positive regulation of apoptotic process	RT		27	13.8	2.8E-16	8.1E-14
	KEGG_PATHWAY	Bladder cancer	RT		16	8.2	3.1E-16	8.9E-15
	KEGG_PATHWAY	Non-small cell lung cancer	RT		19	9.7	6.4E-16	1.6E-14

Beside using DAVID, the list of 197 genes was uploaded into cytoscape cytoHubba, a plug that can measure nodes by their network features to infer their importance in the network, and it can help identify central elements of biological networks. For the analysis it was selected the top 100 proteins of the network and in the Figure 27 shows them in order based on degree.

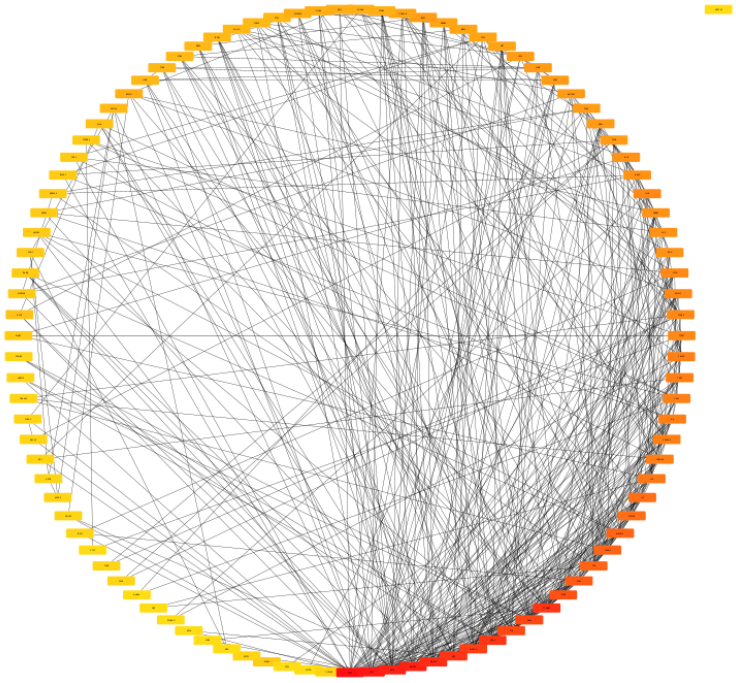
TP53, JUN; MAPK3, AKT1, MAPK1, CTNNB1, HRAS, MYC AND PIK3CA, represented in red are the crucial targets (Figure 27 A) of lung cancer based on the degree and betweenness centrality, in other words, they represent the hub nodes because they are connected to a large size of the

network. Another representation is Figure B with a circular layout it is possible to distinguish more connection with the red boxes and fewer with the genes represented in yellow boxes

Figure 27 Top 100 protein-protein interaction: (A) Grid layout (B) circular layout. (A)



(B)

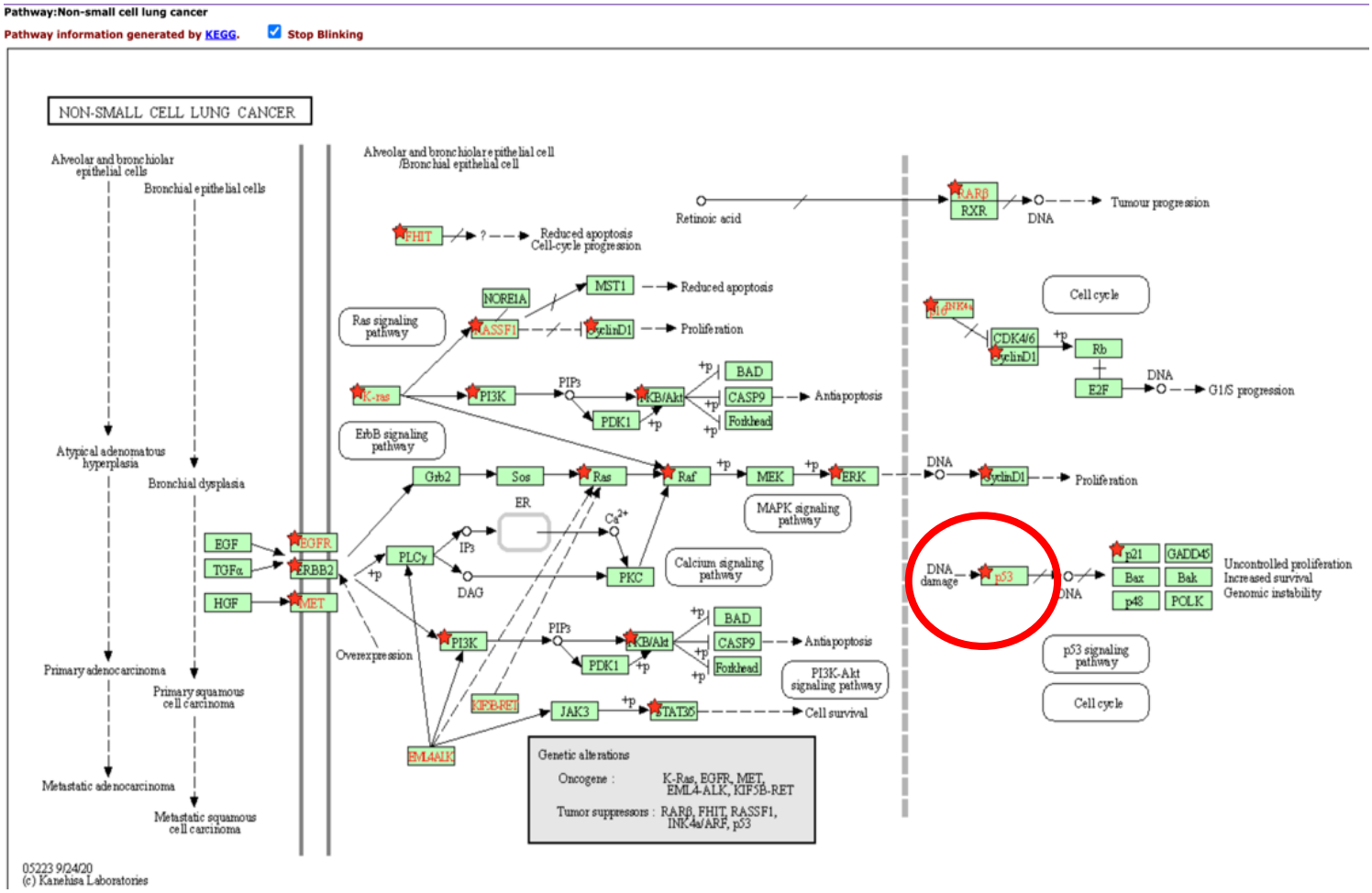


Combining the information from cytoHubba (9 hub genes) and selecting from DAVID the enriched pathways related with lung cancer. Non-small cell lung cancer is the best pathway candidate since is one type of lung disease.

Figure 28 shows the non-small cell lung cancer (NSCLC) pathway, visualized using KEGG, where the boxes with a red star indicate all the genes related to red wine polyphenols and the reactions they are part of. For example, TP53 is interacts with the reaction of DNA damage interaction that lead to uncontrolled proliferation of the cells, as tumor supressor (von Mering et al., 2005). The red circle highlights the TP53 gene, which is a hub node (degree 34) related to 23 red wine polyphenols, such as resveratrol, quercetin, gallic acid, epicagallocatechin gallate, and fisetin, among others (Table 21, Appendix 8.1).

At this point we have reached the objective of finding new possible target using the red wine polyphenols data base. The next step will be look for candidates for future docking studies and potential drug-target interactions.

Figure 28: Kegg pathway of non-small cell lung cancer.



6. DISCUSSION

In this study, we have compiled a set of 144 polyphenols that have been detected in red wine. To build this set, we used different sources, basically Phenol-Explorer, a large database of phenolic compounds, and literature search. A new database has been built from this set of polyphenols, by integrating different variables, using cross reference IDs, from different databases often used in network pharmacology analysis, like PubChem, Chebi, CTD, STITCH, STRING, DAVID and others. The unification between databases was the most challenging step since each of the databases uses different names, and the SMILE or InchiKey were not suitable for search in some of them like in CTD or STITCH. Therefore, the manual cleaning of data was a critical step in the new database.

From the initial 144 polyphenols in the raw database, manual curation identified PubChem ID for 140 of them, which were further used for the analysis in this study (the ones with missing IDs were difficult to analyze, since they were not listed in any of the used databases).

Red wine was the food source with the largest number of different polyphenols among more than 400 foods. This high variety of phenolic compounds is explained by the fact that there are different sources of polyphenols in red wine, like the grape variety, growth conditions, stage of maturity at harvest, the process during the wine making, type of yeast used for fermentation, an many others (Li & Sun, 2019; Tresserra-Rimbau et al., 2018).

More than 60% of the polyphenols in red wine are from the group of flavonoids, which was expected, since this finding is supported from different studies that evaluate polyphenol composition, where anthocyanins, forming the largest subgroup of the flavonoids, are responsible for the quality of the red wine, because they give the major characteristics like color, bitterness, astringency, and chemical stability towards oxidation (Giovinazzo & Grieco, 2015).

Flavonoids, besides being the polyphenol group in red wine with more components, is also the class with the polyphenols found at the highest concentration, being malvidin 3,5-diglucoside the polyphenol with the highest concentration, with an average of 78.62 mg/100 mL, second, delphinidin 3, 5-O-diglucoside with 23.16 mg/100 mL, followed by cyanidin 3,5-O-diglucoside with 20.64 mg/100 mL. This data has been reported using HPLC-ESI-MS/MS (Zhao Q, Duan C, Wang J, 2010). The above mentioned three polyphenols are sub-class anthocyanins diglucosides, found

in Asian wines from grapes *Vitis amurensis* (and not from *Vitis vinifera* which is the principal wine producing plant). *V. amurensis* is a wild Vitis native in north-eastern China, characterized by its resistance to extremely low temperatures of less than -40 °C (Chai et al., 2019; Fennell, 2008).

Since dietary polyphenols, enriched in red wine, are known to have beneficial effects in human health, and moderate consumption of red wine has also been beneficial for some diseases in epidemiological studies, in this work we aimed to find potential protein targets for the polyphenols found in red wine, which could modulate their physiological functions and have an impact in health. Known associations of polyphenols with proteins or genes, which in turn were related to diseases, make it possible to identify polyphenols associated with disease-related proteins. We found associations between red wine polyphenols and a large number of different diseases (69039), including inferred and curated disease-chemical associations, as well as associations of polyphenols with a large number of genes in *H. sapiens* (26674), pathways (13732) and gene ontologies (14583), being resveratrol, quercetin, and epicatechin gallate the ones with more interactions.

The polyphenols with the largest number of associations are actually those involved in a high number of studies. For example, resveratrol has been involved in more than 450 studies (Lucarini et al., 2021), in which the attributed health benefits include cardioprotective effects, chemopreventive activity in different cancers, and a capacity to extend the lifespan of lower organisms (Baur & Sinclair, 2006; Weiskirchen & Weiskirchen, 2016). The beneficial effects have been established only in preclinical models. However, it has been suggested that to obtain the desired health effect it is necessary to take high doses that are not possible to get by drinking wine. For example, if a person intends to ingest 1 g of resveratrol, it will need to drink at least 500 liters (Weiskirchen & Weiskirchen, 2016) due to the fact its concentration in red wine is low (0.27 mg/100mL on average) (Rothwell et al., 2013). Another important fact is that resveratrol is rapidly metabolized in rats, reaching the peak in blood after 10 min of ingestion, but in humans, it is excreted in higher amounts over a 24-h period than either quercetin or catechin, which indicates that maybe it is metabolized to a lower extent. In addition, no microfloral metabolites of resveratrol appear to be produced (Forester & Waterhouse, 2009).

Finally, to probe that the new database is functional, we focused the analysis on lung cancer, which is the second leading cause of cancer-related mortality globally, with an overall five-year survival rate of 22.4% (Lung and Bronchus Cances, n.d.). There are two main types of lung

cancer: the small cell lung cancer (SCLC) that accounts for 15%, and the non-small cell lung cancer (NSCLC), that accounts for the 85% of all lung cancer cases. (Zhou et al., 2019).

The term "lung neoplasm" yielded 197 related genes that interact with red wine polyphenols. This list was enriched with the use of DAVID database that confirm that the genes are indeed enriched in pathways related to cancer. Furthermore, one of the pathways was related to NSCLC.

Using Hubba in cytoscape, we constructed the pharmacology network identifying and prioritizing the influential hub genes in the gene-set. Based on the topology of the network, TP53, JUN; MAPK3, AKT1, MAPK1, CTNNB1, HRAS, MYC AND PIK3CA were identified as the crucial targets for lung cancer, based on the degree and betweenness centrality. They represent the hub nodes, because they are connected to a large part of the network.

Likewise, the pathway for NSCLC was retrieved from KEGG, which showed that TP53, i.e. cellular tumor antigen p53, besides been a hub node in "Lung Neoplasm" it is also associated to NSCLC and acts as a tumor suppressor (von Mering et al., 2005). This finding proposes a potential therapeutic target, since TP53 is related to 23 polyphenols that should be analyzed for bioavailability and docking to continue the studies.

7. CONCLUSION

Network Pharmacology is a suitable tool to discover bioactive compounds for therapeutic targets. Using this analysis, it was possible to build a red wine polyphenol database with 144 polyphenols, identifying curated and inferred association with diseases, genes, pathways and gene ontologies. The database was applied to identify targets for lung cancer based on the genes that are critical according to the enrichment pathway, together with centrality and degree analysis, finding 23 polyphenols affecting TP53 related to NSCLC.

8. FUTURE ESTUDIES

Several future studies are proposed, below described in detail.

Publication of the new database: after the manual filtering and cleaning of the data, a centralized database has been compiled with information about disease, genes, pathways, and gene ontologies of the polyphenols in red wine. We aim to publish the database to contribute to the discovery of new targets for phenolic compounds of red wine.

Regarding metabolites, in this study we did not consider the transformations and metabolism of the polyphenols by microbiome and the human gastrointestinal track because of the lack of data of the metabolic bioproducts. Therefore, it is recommended to study the metabolites that are found in different body fluids, since health benefits of polyphenols might be caused not only by the native molecules but also by their metabolic products.

Docking of the possible targets: the targets for lung cancer described in this study can be used to identify the polyphenols that shows the best binding energy in docking in order to help to develop new potential therapeutic compounds. In further analyses, it would be important to considered bioavailability, biological properties, and health effects, as the latter largely depend on their specific chemical structures.

9. BIBLIOGRAPHY

Baiano, A., Terracone, C., Gambacorta, G., & la Notte, E. (2009). Phenolic Content and Antioxidant Activity of Primitivo Wine: Comparison among Winemaking Technologies. *Journal of Food Science*, *74*(3), C258–C267. https://doi.org/10.1111/J.1750-3841.2009.01101.X

Baur, J. A., & Sinclair, D. A. (2006). Therapeutic potential of resveratrol: The in vivo evidence. In *Nature Reviews Drug Discovery* (Vol. 5, Issue 6, pp. 493–506). https://doi.org/10.1038/nrd2060

Birrell, M. A., McCluskie, K., Wong, S., Donnelly, L. E., Barnes, P. J., & Belvisi, M. G. (2005). Resveratrol, an extract of red wine, inhibits lipopolysaccharide induced airway neutrophilia and inflammatory mediators through an NF-kappaB-independent mechanism. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *19*(7), 1–22. https://doi.org/10.1096/FJ.04-2691FJE

Chai, F., Liu, W., Xiang, Y., Meng, X., Sun, X., Cheng, C., Liu, G., Duan, L., Xin, H., & Li, S. (2019). Comparative metabolic profiling of Vitis amurensis and Vitis vinifera during cold acclimation. *Horticulture Research 2019 6:1*, *6*(1), 1–12. https://doi.org/10.1038/s41438-018-0083-5

Chen, G., Shan, W., Wu, Y., Ren, L., Dong, J., & Ji, Z. (2005). Synthesis and anti-inflammatory activity of resveratrol analogs. *Chemical & Pharmaceutical Bulletin*, *53*(12), 1587–1590. https://doi.org/10.1248/CPB.53.1587

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., & Lin, C. Y. (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, *8*(4), 1–7. https://doi.org/10.1186/1752-0509-8-S4-S11/TABLES/4

Chu, W. M. (2013). Tumor necrosis factor. *Cancer Letters*, *328*(2), 222–225. https://doi.org/10.1016/J.CANLET.2012.10.014

Davis, A. P., Wiegers, T. C., Johnson, R. J., Sciaky, D., Wiegers, J., & Mattingly, C. J. (2022a). Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkac833

Davis, A. P., Wiegers, T. C., Johnson, R. J., Sciaky, D., Wiegers, J., & Mattingly, C. J. (2022b). Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkac833

Diker, N. Y., & Kutluay, V. M. (2021). The evaluation of the antidiabetic effects of red wine polyphenols with the view of in silico prediction methods. *Food Bioscience*, *40*, 100920. https://doi.org/10.1016/j.fbio.2021.100920

do Valle, I. F., Roweth, H. G., Malloy, M. W., Moco, S., Barron, D., Battinelli, E., Loscalzo, J., & Barabási, A. L. (2021). Network medicine framework shows that proximity of polyphenol targets and disease proteins predicts therapeutic effects of polyphenols. *Nature Food*, *2*(3), 143–155. https://doi.org/10.1038/s43016-021-00243-7

Elmali, N., Esenkaya, I., Harma, A., Ertem, K., Turkoz, Y., & Mizrak, B. (2005). Effect of resveratrol in experimental osteoarthritis in rabbits. *Inflammation Research : Official Journal of the European Histamine Research Society ... [et Al.]*, *54*(4), 158–162. https://doi.org/10.1007/S00011-004-1341-6

Fennell, A. (2008). Freezing Tolerance and Injury in Grapevines. *Http://Dx.Doi.Org/10.1300/J411v10n01_09*, *10*(1–2), 201–235. https://doi.org/10.1300/J411V10N01_09

Fernandes, I., Pérez-Gregorio, R., Soares, S., Mateus, N., de Freitas, V., Santos-Buelga, C., & Feliciano, A. S. (2017). Wine Flavonoids in Health and Disease Prevention. *Molecules 2017, Vol. 22, Page 292*, *22*(2), 292. https://doi.org/10.3390/MOLECULES22020292

Flavonoid Nomenclature. (n.d.). Retrieved January 30, 2023, from https://iupac.qmul.ac.uk/flavonoid/

Forester, S. C., & Waterhouse, A. L. (2009). The Journal of Nutrition Supplement: Grapes and Health Metabolites Are Key to Understanding Health Effects of Wine Polyphenolics 1-3. *J. Nutr*, *138*, 1824–1831. https://doi.org/10.3945/jn.109.107664

Giovinazzo, G., & Grieco, F. (2015). Functional Properties of Grape and Wine Polyphenols. *Plant Foods for Human Nutrition*, *70*(4), 454–462. https://doi.org/10.1007/S11130-015-0518-1/FIGURES/2

Hernandez-Ferrer, C., & Gonzalez, J. R. (2018). CTDquerier: A bioconductor R package for comparative toxicogenomics database™ data extraction, visualization and enrichment of environmental and toxicological studies. *Bioinformatics*, *34*(18), 3235–3237. https://doi.org/10.1093/bioinformatics/bty326

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., … D'Eustachio, P. (2020). The reactome pathway knowledgebase. *Nucleic Acids Research*, *48*(D1), D498–D503. https://doi.org/10.1093/NAR/GKZ1031

Jefferson, O. A., Koellhofer, D., Warren, B., & Jefferson, R. (n.d.). *The Lens MetaRecord and LensID: An open identifier system for aggregated metadata and versioning of knowledge artefacts*. https://doi.org/10.31229/OSF.IO/T56YH

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, *51*(D1), D587–D592. https://doi.org/10.1093/NAR/GKAC963

Kashyap, D., Garg, V. K., & Goel, N. (2021). Intrinsic and extrinsic pathways of apoptosis: Role in cancer development and prognosis. *Advances in Protein Chemistry and Structural Biology*, *125*, 73–120. https://doi.org/10.1016/BS.APCSB.2021.01.003

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2022). PubChem 2023 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkac956

King, B. L., Davis, A. P., Rosenstein, M. C., Wiegers, T. C., & Mattingly, C. J. (2012). Ranking Transitive Chemical-Disease Inferences Using Local Network Topology in the Comparative Toxicogenomics Database. *PLoS ONE*, *7*(11), 46524. https://doi.org/10.1371/journal.pone.0046524

Lacroix, S., Klicic Badoux, J., Scott-Boyer, M.-P., Parolo, S., Matone, A., Priami, C., Morine, M. J., Kaput, J., & Moco, S. (n.d.). *A computationally driven analysis of the polyphenol-protein interactome*. https://doi.org/10.1038/s41598-018-20625-5

Li, L., & Sun, B. (2019). Grape and wine polymeric polyphenols: Their importance in enology. *Critical Reviews in Food Science and Nutrition*, *59*(4), 563–579. https://doi.org/10.1080/10408398.2017.1381071

Lopez-Castejon, G., & Brough, D. (2011). Understanding the mechanism of IL-1β secretion. *Cytokine & Growth Factor Reviews*, *22*(4), 189. https://doi.org/10.1016/J.CYTOGFR.2011.10.001

Lucarini, M., Durazzo, A., Lombardi-Boccia, G., Souto, E. B., Cecchini, F., & Santini, A. (2021). Wine polyphenols and health: Quantitative research literature analysis. *Applied Sciences (Switzerland)*, *11*(11). https://doi.org/10.3390/app11114762

Lung and Bronchus Cancer — Cancer Stat Facts. (n.d.). Retrieved January 27, 2023, from https://seer.cancer.gov/statfacts/html/lungb.html

Markoski, M. M., Garavaglia, J., Oliveira, A., Olivaes, J., & Marcadenti, A. (2016). Molecular Properties of Red Wine Compounds and Cardiometabolic Benefits. *Nutrition and Metabolic Insights*, *9*, 51. https://doi.org/10.4137/NMI.S32909

Namiki, K., Wongsirisin, P., Yokoyama, S., Sato, M., Rawangkan, A., Sakai, R., Iida, K., & Suganuma, M. (2020). (−)-Epigallocatechin gallate inhibits stemness and tumourigenicity stimulated by AXL receptor tyrosine kinase in human lung cancer cells. *Scientific Reports 2020 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-59281-z

Neveu, V., Perez-Jimé Nez, J., Vos, F., Crespy, V., du Chaffaut, L., Mennen, L., Knox, C., Eisner, R., Cruz, J., Wishart, D., & Scalbert, A. (n.d.). *Phenol-Explorer: an online comprehensive database on polyphenol contents in foods*. https://doi.org/10.1093/database/bap024

Nicholson, K. M., & Anderson, N. G. (2002). The protein kinase B/Akt signalling pathway in human malignancy. *Cellular Signalling*, *14*(5), 381–395. https://doi.org/10.1016/S0898-6568(01)00271-6

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., & Demchak, B. (2019). Cytoscape Automation: Empowering workflow-based network analysis. *Genome Biology*, *20*(1). https://doi.org/10.1186/S13059-019-1758-4

Panche, A. N., Diwan, A. D., & Chandra, S. R. (2016). Flavonoids: an overview. *Journal of Nutritional Science*, *5*, 1–15. https://doi.org/10.1017/jns.2016.41

Pérez-Jiménez, J., Neveu, V., Vos, F., & Scalbert, A. (2010). Identification of the 100 richest dietary sources of polyphenols: an application of the Phenol-Explorer database. *European Journal of Clinical Nutrition*, *64*, 112–120. https://doi.org/10.1038/ejcn.2010.221

Quideau, S., Deffieux, D., Douat-Casassus, C., & Pouységu, L. (2011). Plant polyphenols: Chemical properties, biological activities, and synthesis. In *Angewandte Chemie - International Edition* (Vol. 50, Issue 3, pp. 586–621). https://doi.org/10.1002/anie.201000044

Rönnstrand, L. (2004). Signal transduction via the stem cell factor receptor/c-Kit. *Cellular and Molecular Life Sciences*, *61*(19–20), 2535–2548. https://doi.org/10.1007/S00018-004-4189-6

Rothwell, J. A., Perez-Jimenez, J., & Neveu, V. (2013). Database Update Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content0: a major update of the Phenol-Explorer database to incorpo-rate data on the effects of food processing on polyphenol content. *Phenol-Explorer*, *3*. https://doi.org/10.1093/database/bat070

Singla, R. K., Dubey, A. K., Garg, A., Sharma, R. K., Fiorino, M., Ameen, S. M., Haddad, M. A., & Al-Hiary, M. (2019). Natural Polyphenols: Chemical Classification, Definition of Classes,

Subcategories, and Structures. *Journal of AOAC INTERNATIONAL*, *102*(5), 1397–1400. https://doi.org/10.1093/JAOAC/102.5.1397

Style and Usage for Organic Chemistry. (2009). *The Manual of Scientific Style*, 399–425. https://doi.org/10.1016/B978-012373980-3.50014-9

Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., & Kuhn, M. (2016). STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Research*, *44*. https://doi.org/10.1093/nar/gkv1277

Tresserra-Rimbau, A., Lamuela-Raventos, R. M., & Moreno, J. J. (2018). Polyphenols, food and pharma. Current knowledge and directions for future research. *Biochemical Pharmacology*, *156*, 186–195. https://doi.org/10.1016/J.BCP.2018.07.050

Visioli, F., Panaite, S. A., & Tomé-Carneiro, J. (2020). Wine's Phenolic Compounds and Health: A Pythagorean View. *Molecules 2020, Vol. 25, Page 4105*, *25*(18), 4105. https://doi.org/10.3390/MOLECULES25184105

von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., & Bork, P. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Research*, *33*(Database issue). https://doi.org/10.1093/NAR/GKI005

Vong, C. I., Rathinasabapathy, T., Moncada, M., & Komarnytsky, S. (2022). All Polyphenols Are Not Created Equal: Exploring the Diversity of Phenolic Metabolites. In *Journal of Agricultural and Food Chemistry* (Vol. 70, Issue 7, pp. 2077–2091). American Chemical Society. https://doi.org/10.1021/acs.jafc.1c07179

Weiskirchen, S., & Weiskirchen, R. (2016). Resveratrol: How Much Wine Do You Have to Drink to Stay Healthy? *Advances in Nutrition*, *7*(4), 706–718. https://doi.org/10.3945/AN.115.011627

Xingyu, Z., Peijie, M., Dan, P., Youg, W., Daojun, W., Xinzheng, C., Xijun, Z., & Yangrong, S. (2016). Quercetin suppresses lung cancer growth by targeting Aurora B kinase. *Cancer Medicine*, *5*(11), 3156–3165. https://doi.org/10.1002/CAM4.891

Xinyi, X., Leping, L., Xueshuai, C., Xi, L., Sujuan, P., & Guomin, Z. (2022). Network pharmacology and molecular docking analysis on molecular targets and mechanism prediction of Huanglian Jiedu Decoction in the treatment of COVID-19. *Digital Chinese Medicine*, *5*(1), 18–32. https://doi.org/10.1016/J.DCMED.2022.03.003

Yousef, M., Vlachogiannis, I. A., & Tsiani, E. (2017). Effects of Resveratrol against Lung Cancer: In Vitro and In Vivo Studies. *Nutrients 2017, Vol. 9, Page 1231*, *9*(11), 1231. https://doi.org/10.3390/NU9111231

Zhou, Q., Pan, H., & Li, J. (2019). Molecular Insights into Potential Contributions of Natural Polyphenols to Lung Cancer Treatment. *Cancers 2019, Vol. 11, Page 1565*, *11*(10), 1565. https://doi.org/10.3390/CANCERS11101565

10. Appendix

10.1 Index of tables

10.2 Index of Figures

10.3 Web page and Rmarkdown code

For further details of the methodology, a step by step guide can be found online: https://model3dbio.csic.es/poli_database_redwine