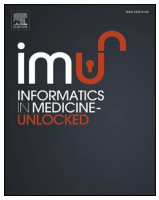
Informatics in Medicine Unlocked 36 (2023) 101163

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu



Molecular network analysis of hormonal contraceptives side effects via database integration

Manuela Petti [a,*], Caterina Alfano [b], Lorenzo Farina [a]

[a] Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto, 25, 00185, Rome, Italy
[b] Department of Experimental Medicine, Sapienza University of Rome, Viale Regina Elena, 324, 00161, Rome, Italy

ARTICLE INFO

Keywords:
Hormonal contraceptives
Multipartite graph
Network medicine
Side effect module
Precision medicine

ABSTRACT

Hormonal contraceptives (HCs) have been shown to be safe and effective when used correctly and consistently, however, as other classes of drugs, they are also associated with adverse health outcomes. In this study, we aim to explain the occurrence of common and unexpected HCs side effects (SEs) integrating drug-target, drug-SE and protein-protein interaction (PPI) public databases. We created a tripartite network that includes three types of vertices: SEs, drugs, and targets. The three layers are linked by means of the inter-layer associations drug-target and drug-SE, whereas only the target layer is characterized also by intra-layer links (PPIs). We exploited the drug-mediated association SE-target to identify the side effect modules defined as a network connected component composed of target proteins plus the proteins needed to connect them. We found that module proteins are associated with diseases/phenotypes and/or KEGG pathways related to the SEs. In particular, in many cases, targets are not enriched in SE features, whereas investigating their neighborhood (here defined as the proteins that allow the targets' connection) we found SE-related pathways. These results show that HCs action can perturb the targets' neighborhood inducing unwanted reaction and that the proposed approach can help to understand how, and through which molecular mechanisms, side effects can occur. The approach is general in its nature: it can be applied to other drugs categories providing a support in identifying a subject-specific therapy that takes into account comorbidities and lifestyle to reduce or avoid the most undesired side effects.

1. Introduction

To date, hormonal contraception is the most used contraceptive method: hormonal contraceptives (HCs) are made of synthetic forms of the hormones progesterone and oestrogen and prevent ovulation by maintaining more consistent hormone levels. There are two main types of HCs formulations: combined methods which contain both an oestrogen and a progestin, and progestogen-only methods which contain only progestins. Combined hormonal contraceptives were developed to prevent ovulation by suppressing the release of gonadotropins to avoid an increase in estradiol levels. Progestogen-only methods, instead, rely more heavily on changes in cervical mucus, which reduces sperm viability and penetration.

Overall, these drugs have been shown to be safe and effective when used correctly and consistently, however they are also associated with adverse health outcomes. For HCs this is of particular interest: in fact, side effects are one of the main causes of drug discontinuation and oral contraceptive misuse and suspension, leading to more than three quarters of a million unintended pregnancies among young U.S. women each year [1]. For example, progesterone-only contraceptive methods, such as implants and hormonal intrauterine devices, tend to trigger or worsen many dermatological conditions, including acne, hirsutism, alopecia, and even rosacea [2], which have a strong impact on a social level and are hence highly unwanted by women. There are also some more serious risks that scare women off from using hormonal contraceptives; for example, it has been found that combined oral contraceptives use was associated with approximately two to four times the risk of stroke compared with non-use, especially among women with migraine with aura [3]. Furthermore, other common side effects are depression [4], cardiovascular disease [5], alteration of glucose level [6] and more.

In general, the drug development process is made of many steps, starting from lab analysis, and going to animal testing, then human testing and finally the reviews and approvals followed by a constant post-market safety monitoring [7]. This process is extremely slow, costly and with a low success rate: approximately 90% of new drug researches do not make it past the early development and toxicity testing, and many

* Corresponding author.
E-mail address: manuela.petti@uniroma1.it (M. Petti).

https://doi.org/10.1016/j.imu.2023.101163
Received 11 October 2022; Received in revised form 2 January 2023; Accepted 3 January 2023
Available online 4 January 2023

of the few drugs that make it to clinical trials fail because of side-effects or adverse events. Moreover, also referring to the drugs on the market, the mechanisms of many of them are still scarcely outlined, and unexpected and severe side effects can occur. This difficulty in understanding the mechanisms behind adverse drug reactions (ADRs) is a consequence of the difficulty of predicting the behavior of a drug once it enters the human body: even if the drug was specifically chemically designed to match a target protein, it might still match and interact with an off-target (a protein with a similar structure to the target) or it might also interact with other drugs or with nutrients (like food and alcohol).

Recently, computational approaches have been developed to investigate the mechanisms behind side effects by trying to connect the molecular scale (drug-protein interactions) and the phenotypic scale (side effects). The biggest trend that emerged is to exploit freely available knowledge- and data-bases, network analysis and multidisciplinary skills. For example, Kuhn et al. "integrated phenotypic data obtained during clinical trials with known drug–target relations to identify overrepresented protein–side effect combinations" [8]: the authors used the sets of predicted targets to explain the majority of the studied side effects, however their approach can only make predictions for proteins that are the targets of a certain number of drugs. Lee et al. aimed to "automatically discover the relationship between biological processes and side effects by building a multi-level network of drug-biological processes influenced by the association of targets with side effects". Their hypothesis is that if some drugs cause the same side effect, their biological process is thought to be possibly related to their side effect [9]. In another study, the central idea was to apply a large-scale analysis to identify overrepresented ADR-pathway combinations through merging clinical phenotypic data, biological pathway data, and drug-target relations. They found that frequent ADRs were associated with more pathways than rare and infrequent ADRs, and that perturbing a certain pathway can cause changes in multiple organs, rather than in one specific organ [10]. Furthermore, other approaches [11,12] suggested the use of the concept of network module. This concept comes from the definition of modularity, a measure of the strength of division of a network into modules: thus, a module is a set of nodes densely connected to each other and with sparse connections with nodes in different modules. A first application of this concept was proposed by Chen et al. [12]: their aim was to build and study an ADR-protein network to identify ADR-ADR associations and predict new ADR-related proteins. In particular, they applied an agglomerative clustering method and identified highly modular sets composed of both ADRs and ADR-related proteins [14]. Later, Guney [11] defined and analyzed side-effect modules with the aim to predict side effects not present in SIDER (a database containing information on marketed medicines and their recorded adverse drug reactions [13]. Using proximity measures, he found that drug targets are closer in the interactome to the proteins inducing the known side effects.

From the above introductory discussion, the understanding and the prediction of side effects is still an open problem. In particular, all the proposed methodologies consider all the types of drugs in the same model: this approach can help in understanding relations between diseases and similarities between different therapies providing potential insights for drug repositioning purposes. However, coming to a conclusion about a specific type of therapy based on an aspecific investigation, cannot help in the selection of the most appropriate drug to avoid or limit specific side effects based on single-patient needs.

In this study, we aim to characterize frequent and unexpected HCs side effects at the molecular level by integrating databases (drug-target associations, drug-side effect associations, protein-protein interactions) and investigating the role of the drug target network neighborhood. To reach this goal, we first identify the set of target proteins eliciting each side effect, and then we project these proteins onto the human interactome: if they are not a connected component, we build the side effect module adding the proteins needed to connect them. Lastly, for each side effect, we used the enrichment analysis tool Enrichr [14–16] for checking if the associated module is enriched in diseases/phenotypes and/or KEGG pathways that explain the side effect.

2. Materials and methods

2.1. Drugs and targets

To collect data about drugs and their targets we used Drugbank [17]. Our version of the database dates to March 24, 2021: at that moment 14315 drugs were available. To select the specific set of drugs (i.e., hormonal contraceptives), we used the MeSH Id associated with each drug. Inside the category "Contraceptive Agents", we considered everything except spermicides, hormonal contraceptives for males or drugs which had only an abortive function (while keeping the ones which acted both as a contraceptive and abortive medication). Finally, we decided to focus exclusively on drugs labeled as approved or investigational-approved.

In the end, we obtained 24 drugs which matched all these requirements. However, only for 21 of them there was information about their targets on Drugbank. Moreover, 3 of them (estradiol benzoate, estradiol cypionate and estradiol valerate) were considered as one drug because they are made with the same molecule and differ only in the attached fatty acids, which allow different properties during administration; in fact, these drugs have the exact same targets. So, the final set of analyzed drugs and targets was made of 19 drugs and 21 targets (Table 1).

We also checked the available information about drugs' other protein interactors: enzymes, transporters and carries. The enzymes of this drug category, as well as the transporters, generally belong to the same family. Indeed, all the drugs exploit Cytochromes P450 (*CYPs*) family for the enzymes (11 out of 18 exclusively *CYPs,* while the others mainly add proteins of the UGT family to other CYP enzymes), and Solute carrier (SLC) family and ATP-binding cassette (ABC) family for the transporters. More generally, these protein families are commonly used by many medications of different categories, for example CYP enzymes metabolize 90 percent of drugs [18]. For this similarity between all the studied drugs in terms of enzymes and transporters, we decided to exclude these categories of drugs interactors to avoid not discriminating specific protein-side effect associations. Lastly, the carriers used by HCs are ALB and SHBG: these proteins are included in the study because they are also targets.

Table 1
Set of analyzed hormonal contraceptives and their targets.

Drug Name	Target
Etonogestrel	PGR
Desogestrel	PGR ESR1
Megestrol acetate	PGR NR3C1
Levonorgestrel	PGR ESR1 NR3C1 SRD5A1 AR SHBG
Progesterone	PGR ESR1 NR3C1 AR SHBG NR3C2 CYP17A1 OPRK1 ORM1 ESR2
Norethisterone	PGR NR3C1 AR
Estradiol	ESR1 ESR2 NR1I2 CHRNA4 GPER1 MT-ATP6 BECN1
Ethynodiol diacetate	PGR ESR1
Norgestimate	PGR ESR1 AR
Ethinylestradiol	ESR1 NR1I2
Mestranol	ESR1
Drospirenone	PGR NR3C1 AR NR3C2
Estrone sulfate	ESR1 ESR2
Norelgestromin	PGR AR ALB
Hydroxyprogesterone caproate	PGR
Norethynodrel	ESR1 AR SHBG
Norgestrel	PGR SRD5A1 AR
Gestrinone	PGR ESR1 NR3C1 AR SHBG GNRHR
Estradiol valerate	ESR1 ESR2 NR1I2 CHRNA4 GPER1 MT-ATP6 BECN1 NCOA2 HSD17B2 ESRRG

2.2. PPI network

The Protein-Protein Interaction (PPI) network was retrieved from the literature [19]. It contains a total of 16470 nodes (proteins) and 233957 edges (protein-protein interactions). Amongst the 21 considered targets, only one of them (HSD17B2) was not present in the network and was therefore ignored during the analysis.

2.3. Side effects

Side effects (SEs) were gathered from:

- Offside Database (downloaded on 12 April 2021) [20].
- Sider Database (downloaded on 7 April 2021) [13].
- FDA Labels (FDA sources consulted in April 2021)
- RxList, an online medical resource (consulted in April 2021) [21].
- NDrugs, an online resource about generic drugs (consulted in April 2021) [22].
- Wikipedia (consulted in April 2021)

As none of these sources contained exhaustive information on all of the 19 considered drugs, we have manually integrated the information applying the same selection criteria for all of them, to obtain consistent and coherent data.

For each drug, we examined each of the listed resources, and collected the related side effects only when information about the side effect frequency and the placebo effect frequency were available or could be calculated from the data provided. Regarding the frequency, it is worth noting that the drug-side effect associations do hold a degree of subjectivity. In fact, drug safety surveillance strongly relies on spontaneous reporting of adverse reactions/side effects. This system is at times affected by underreporting, which might be due to many causes like lack of knowledge about pharmacovigilance, lack of time, the belief that only serious or previously unknown ADRs should be reported and more [23,24]. Moreover, there is not a standard method to present toxicity data; the FDA has released guides seeking to standardize adverse effects reporting, but they are more subjective than concrete [25]. On the other hand, over-reported drug-SE associations can happen as a consequence of topics of broad and current interest (autism, vaccines) and are also known to be frequent in mental-health patients [26]. The above-mentioned biases in drug-SE associations' frequency can affect the selection of SE of interest, however in our work, they do not have an impact on the method used to analyze them.

In detail, For the textual sources (FDA labels, RxList, NDrugs, Wikipedia), we used the frequency definitions given by the Council for International Organizations of Medical Sciences (CIOMS) as reported by WHO: following this guide, we only selected side effects that were either classified as common or as very common. As for the placebo effect, we assumed that the information present on the labels (and therefore also on the websites that reported it) was already taking it into account, as stated on the FDA website [27]. On the other hand, for both Offsides and Sider databases, we set specific thresholds to the available metrics. In particular, in Offsides, we considered the "mean_reporting_frequency" (mrf) and the "Proportional Reporting Value" (PRR), setting the following thresholds: mrf >0.01 (to consider side effects that are at least common) and PRR >1 (to consider only the side effect for which the real frequency is greater than the placebo one). Analogously, for the side effects selection from Sider, we applied the criteria: the side effect frequency had to be greater than 0.01 while also being greater than the placebo frequency (we computed the PRR ourselves by exploiting the available data).

As a result of the side effects collection process, we excluded the drug *Norethynodrel,* because the needed information was not available on any of the consulted sources. Furthermore, there are two special cases: for two of 18 remaining drugs, there wasn't any information or commercial product with the molecule on its own; by looking at the drugs' labels we found that they are (generally/mostly) sold as a combined medication with another molecule. Hence, from now on, we will refer to "ethynodiol diacetate" and "norelgestromin" as "ethynodiol diacetate + ethinylestradiol" and "norelgestromin + ethinylestradiol".

Finally, similar or too specific side effects were grouped into broader categories and renamed appropriately; for example, breast tenderness, breast enlargement or swelling, breast pain etc. were all considered as breast discomfort (see Table S1 for all the details). In the end, for the analyses that follow, we considered only unexpected side effects, the less accepted by the patients and often unexplained ones. We considered as expected those SEs that affect the reproductive system (e.g. breast discomfort and changes in menstrual bleeding patterns) and thus we finally obtained 56 side effects.

2.4. Side effects modules

To try to explain the side effects at the molecular level, we first identified which are the proteins related to a particular side effect by integrating the different types of relations described above: protein-protein interactions, drug-target associations, drug-side effect associations. Thus, we built a multipartite graph: in our case we have three layers (target-drug-side effect) as shown in Fig. 1.

To associate the side effects with proteins, we substituted the direct link between the two pairs "target-drug" and "drug-side effect", with the indirect and drug-mediated relation "target-side effect": this is equivalent to hiding the middle layer (the drug layer) inside the tripartite graph.

Doing so, for each side effect, we obtained the set of targets associated with it on at least one occasion. At this point, we projected these target proteins onto the PPI network and, in the cases they are not a connected compontent, we further investigated their neighborhood. To identify the targets' neighborhood, we exploited the concept of disease module proposed by Wang and Loscalzo [28] defining the side effect module as a network connected component composed of target proteins plus the proteins needed to connect them. Each module is thus obtained by starting from the target proteins and expanding the module with an iterative process that leads to the addition of proteins that connect the disconnected targets. This process conforms to the Seed Connector Algorithm (SCA) proposed by Wang and Loscalzo [28], differing in the fact that we will not stop the process until we obtain a connected component and, in doing so, we do not discriminate between paths of the same length needed to connect the remaining targets. In the rest of the work, we refer to this process using the definition of Target Connector Algorithm (TCA).

After identifying the SE modules, for each side effect, we performed the enrichment analysis of the module through EnrichR [14–16]. To evaluate the perturbation induced by targeting drug proteins, we focused also selectively on the connector proteins (the secondary targets) of each module performing their enrichment analysis and confronting the results with the current available literature.

3. Results and discussion

3.1. The bipartite networks

We used the data gathered from Drugbank to create a bipartite drug-target network with a total of 40 nodes belonging to the two different groups; this network has 71 edges, leading to a 9% density. The network is shown in Fig. 2 where drugs are colored in orange and targets in green, while the size of a node is directly proportional to the number of links (i.e. drug-target associations) it has. Looking at the nodes' distribution, it is evident that targets mostly have a degree smaller than 3 with a maximum value of 13, while drugs can reach a maximum of 10 and present less cases that have only one connection (i.e. one target).

With the same process we built the drug-side effect network (Fig. S1). This network is made of 18 drugs and 82 side effects and has a density of

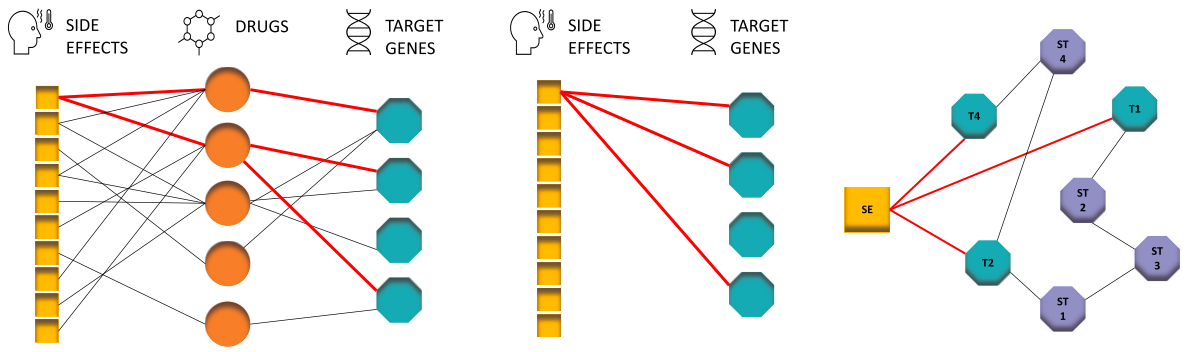


Fig. 1. Toy representation of the proposed procedure: the tripartite network including the three types of vertices on the left; SE-target bipartite network obtained substituting the direct link between the two pairs "target-drug" and "drug-SE" with the indirect and drug-mediated relation "target-SE" in the middle; side effect module composed of target proteins (in teal) and secondary target (in lilac) on the right.

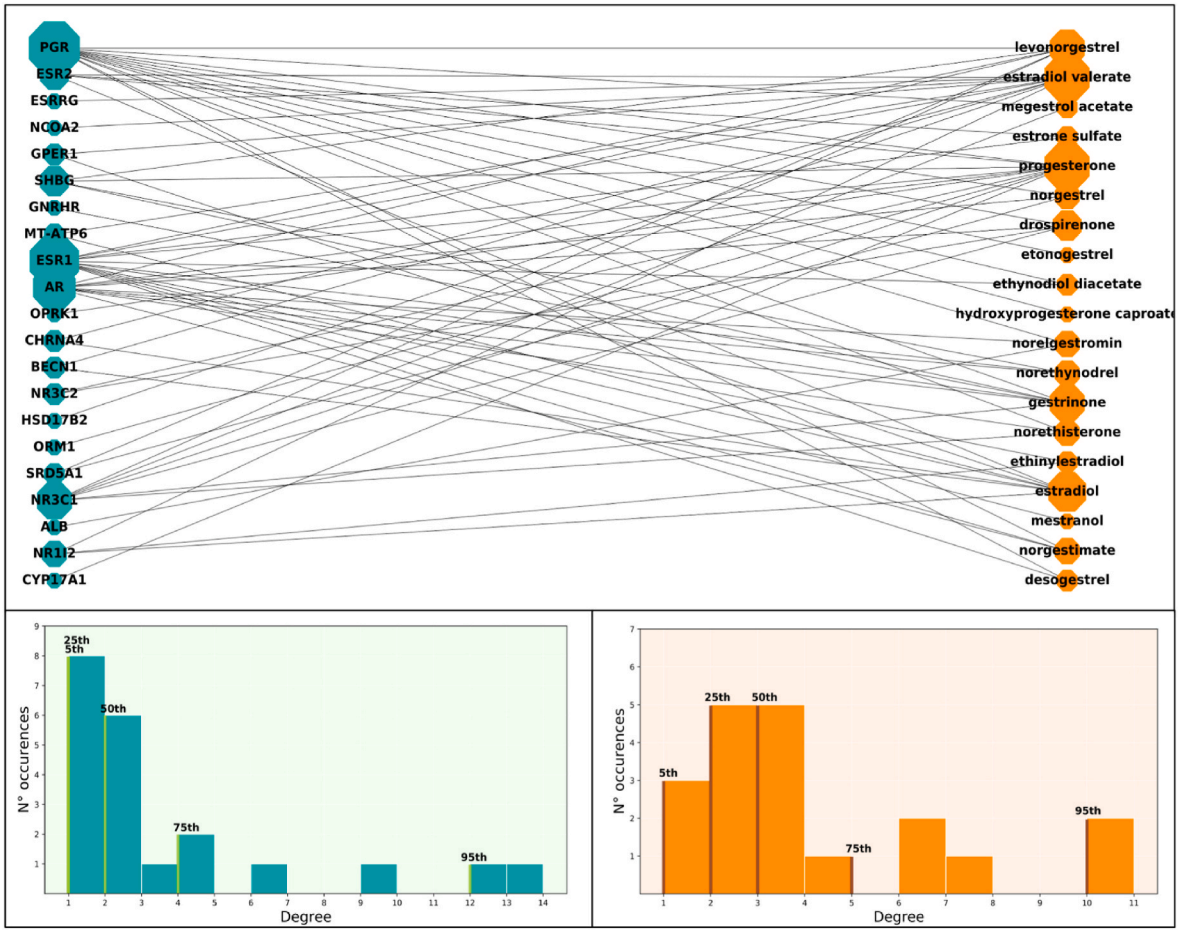


Fig. 2. The drug-target bipartite network: targets are on the left side colored in teal, while drugs are on the right colored in orange. The bigger the size of a node, the more interactions it has. The degree distribution is shown for the two sets of nodes: targets and drugs in the lower left and right panels respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

7%; half of the drugs have less than 18 side effects, but no drug has less than 3. On the other hand, many side effects are associated with only one drug, while few of them are associated with the majority of the drugs (e.g. headache, weight change).

3.2. Tripartite network

The final tripartite network is made of 18 drugs, 21 targets and 82 side effects, and it is the results of the union between the drug-target network and in the drug-side effects network; it is structured in 3 layers, with drugs at the center allowing the indirect connection of targets and side effects. The tripartite network is shown in Fig. 3 and can be found in the supplementary material in the form of an edge list.

The target layer is the only one characterized also by intra-layer links (protein-protein interactions): most of the target proteins didn't have any kind of interaction among themselves, except for a small group of 7 targets that make up a connected component (see Fig. 3). Amongst these seven proteins, there are some of the main targets of hormonal contraceptives like PGR and ESR1.

Approximately half of the considered drugs only target proteins

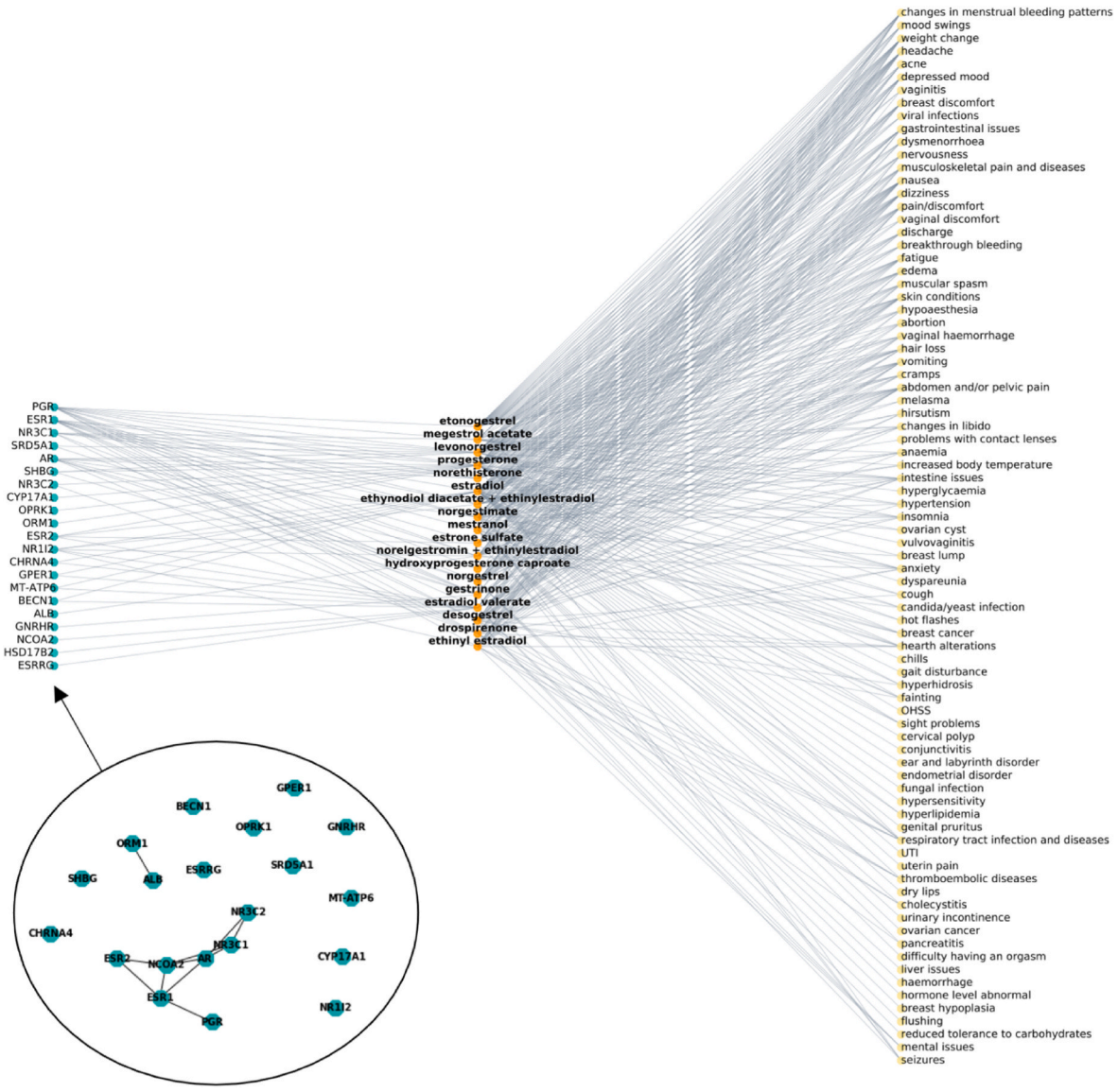


Fig. 3. The tripartite network: targets are on the left side colored in teal, drugs are in the middle layer colored in orange and side effects are on the right colored in yellow. The target layer is the only one characterized also by intra-layer links (protein-protein interactions) shown in the circle at the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

belonging to this connected component (class D1, see Table 2), while the other ten drugs have more diversified targets (class D2).

Starting from this categorization of the drugs, we identified the side effects that only happen when using drugs belonging to D1 and thus linked only to proteins that already make up a connected component. Given the impossibility to construct a module and the small number of targets involved to perform the enrichment analysis, we decided to discard these side effects which are: increased body temperature, urinary incontinence, dry lips and pancreatitis (for example urinary incontinence is associated only with ESR1).

Table 2
Drugs classification.

D1	Etonogestrel, Desogestrel, Megestrol acetate, Norethisterone, Norgestimate, Mestranol, Drospirenone, Estrone sulfate, Hydroxyprogesterone caproate
D2	Levonorgestrel, Progesterone, Estradiol, Ethynodiol diacetate + Ethinylestradiol, Norelgestromin + Ethinylestradiol, Norgestrel, Gestrinone, Estradiol valerate, Norethynodrel

3.3. Module-based explanation of the HCs side effects

As described in the previous sections, we focused on the frequent and unexpected side effects (N = 52) of hormonal contraceptives identifying their associated drug target proteins. In our study, we do not assume the unique hypothesis that a side effect can be elicited by the associated targets, but we also assume that the drug's action in targeting specific proteins may induce a perturbation in the targets' neighborhood. Based on these hypotheses, to try to understand the molecular mechanisms of the HCs side effects, we exploited the definition of the disease module [29,30]. This module definition goes beyond the concept of topological module (a locally dense network neighborhood) because it is based on the hypothesis of its overlap with a connected component of genes involved in the same disease/phenotype and with functional modules (aggregation of genes with similar or related functions) [31]. Several module-based algorithms have been proposed to identify the disease modules [28,32–34] exploiting topological and/or functional features of known disease-associated genes and, among them, we decided to exploit the method proposed in Ref. [28]. This choice is based on the fact that for the purposes of the study, we do not aim at the complete

topological/functional characterization of the target proteins to find other similar ones, rather we want to investigate if a side effect may be induced as a perturbation in the neighborhood common to and linking not connected targets. To do this we have proposed the Target Connector Algorithm (TCA), a modified version of SCA which fits the purpose of the study of side effects, and its application to the HCs adverse reactions identified 23 modules.

Each identified module can be associated with one or more side effects. This is due to the fact that some drugs use many of the available therapeutic targets, so side effects associated with many drugs end up being linked to the same targets and, as a consequence, having the same side effect module. In particular, thirteen out of 23 modules are uniquely associated with a side effect, the other 10 are instead associated with more side effects. Fig. 4-A shows the composition of each module: their size varies in the range 3–46 nodes and in more than half of the cases (13 out of 23 modules), the number of secondary targets (i.e. the proteins connecting drug targets) is greater than the number of the targets. Interestingly, the size of the identified modules is not correlated with the number of associated side effects (see Fig. 4-B). In fact, two of the largest modules (both composed of 44 proteins) represent two opposite situations: one of them (module 2 in Fig. 4) is related to a single side effect, while on the other (module 3 in Fig. 4), 7 different side effects are mapped.

Below, we provide a detailed description of the module-based explanation of a selected set of HCs side effects. The selection includes common side effects of many HCs (e.g. depressed mood occurs as a side effect of 10 different drugs) and side effects related to greater concern in the use of hormonal contraception (e.g weight change), that usually also cause the discontinuation of the drug.

3.3.1. Depressed mood

Depressed mood occurs as a side effect of 10 different drugs. The enrichment analysis performed with EnrichR (category: Diseases/Drugs, section: DisGeNET) revealed that target proteins are strongly enriched in "Mental Depression" (adjusted p-value $= 6.093 \times 10^{-11}$) and "Depressive Disorders" (adjusted p-value $= 2.201 \times 10^{-10}$), however they are not significantly involved in pathways related to depression (EnrichR, category: Pathways, section: KEGG 2021 Human). On the other hand, by looking at the 28 secondary targets, we found a significant enrichment in the KEGG pathways: "Oxidative phosphorylation" (adjusted p-value $= 2.41 \times 10^{-5}$) and "Pathways of neurodegeneration" (adjusted p-value $= 4.83 \times 10^{-4}$) which have been studied for their role in the development of depressive symptoms [35,36].

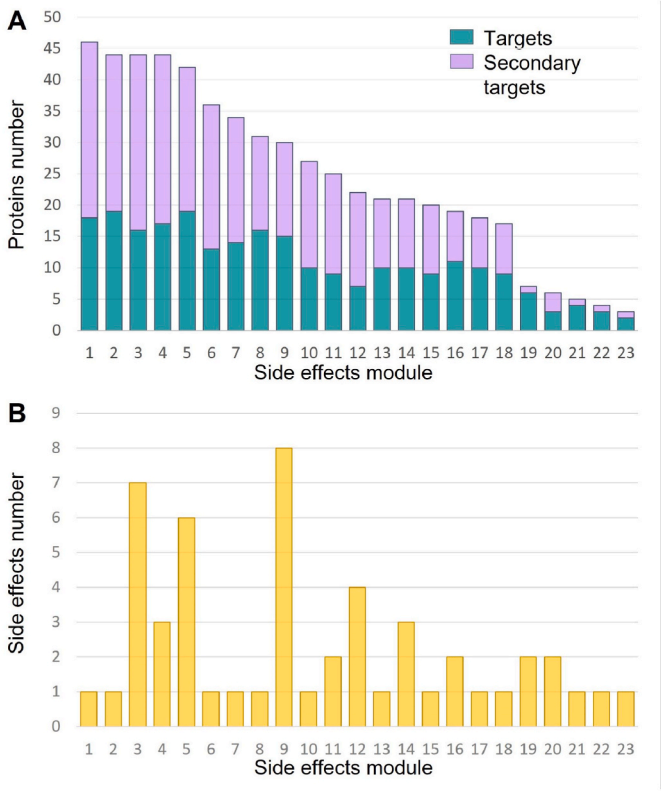


Fig. 4. Bar diagrams showing the number of proteins that make up each module distinguishing between target and secondary proteins (panel A) and the number of associated side effects (panel B).

3.3.2. Nervousness

Nervousness is a quality of feeling anxious, worried, or alarmed with respect to a normal and calmer state of mind. This side effect occurs with 7 of the investigated drugs. A first explanation can be found in the targets: many of them are in fact associated with "Mood Disorders" (EnrichR, category: Diseases/Drugs, section: DisGeNET; adjusted p-value $= 3.71 \times 10^{-6}$) and "Social Stress" (EnrichR, category: Diseases/Drugs, section: DisGeNET; adjusted p-value $= 4.31 \times 10^{-4}$). However, it is worth noting that only the secondary targets are significantly enriched in the "Retrograde endocannabinoid signaling pathway" (adjusted p-value $= 5.13 \times 10^{-5}$). This pathway is relevant when studying a nervous or stressed state of mind because "The endocannabinoid (eCB) system has emerged as a central integrator linking the perception of external and internal stimuli to distinct neurophysiological and behavioural outcomes (such as fear reaction, anxiety and stress-coping), thus allowing an organism to adapt to its changing environment" [37].

3.3.3. Musculoskeletal pain and diseases

Musculoskeletal disorders are injuries or disorders of the muscles, nerves, tendons, joints, cartilage, and spinal discs. Again, the targets were found to be involved in some musculoskeletal diseases such as "Degenerative polyarthritis" (adjusted p-value $= 5.82 \times 10^{-5}$), "Rheumatoid Arthritis" (adjusted p-value $= 2.55 \times 10^{-4}$), "Arthritis" (adjusted p-value $= 6.26 \times 10^{-4}$), "Osteoarthritis of the hand" (adjusted p-value $= 2.03 \times 10^{-3}$) and "Osteoarthritis, Knee" (adjusted p-value $= 8.29 \times 10^{-3}$). The secondary targets reinforce the relationship between this module and musculoskeletal disorders by showing a significant enrichment in the "Relaxin signaling pathway" (adjusted p-value $= 6.26 \times 10^{-3}$); relaxin is a hormone that "exerts its regulatory effect on the musculoskeletal and other systems through binding to its receptor in various tissues, mediated by different signaling pathways. Relaxin alters the properties of cartilage and tendon by activating collagenase. This hormone is also involved in bone remodeling and healing of injured ligaments and skeletal muscle" [38].

3.3.4. Weight change

Weight change (and in particular weight gain) is commonly cited as a side effect of hormonal contraception, typically the cause of discontinuation or reluctance to initiate. The potential association between hormonal contraception and changes in weight is still an open debate because a causal relationship between them has not been established [39,40]. Our analysis reveals that the module identified for this side effects (42 proteins including 19 targets and 23 secondary targets) is strongly enriched (category: Diseases/Drugs, section: DisGeNET) in "Obesity" (adjusted p-value $= 5.508 \times 10^{-9}$) and "overweight" (adjusted p-value $= 2.227 \times 10^{-8}$). Both the targets and the secondary targets contribute to these significant enrichments; however, it is worth noting that only the secondary targets are characterized by a significant association with "Obesity" using OMIM database (adjusted p-value $= 5.764 \times 10^{-3}$). Indeed, among them the presence of the POMC gene stands out as it is a key mediator of satiety and gene mutations are associated with obesity [41]. Furthermore, the new proteins associated to weight change are enriched in thermogenesis (adjusted p-value $= 5.764 \times 10^{-3}$) and oxidative phosphorylation (adjusted p-value $= 6.296$

$\times 10^{-3}$), both these pathways have been studied in body weight change and maintenance [42].

3.3.5. Ear and labyrinth disorders

This side effect appears as an adverse reaction only to Estradiol. The module returned by TCA is the one for which the ratio between secondary targets and targets is the highest (14/7). By looking at enrichment in KEGG pathways and DisGeNET disease-gene associations, for the seven targets linked to this side effect, we could not find an explanation for it. The secondary targets are instead significantly enriched in "Ras signaling pathway", "Calcium signaling pathway" and "Regulation of actin cytoskeleton" with an adjusted p-value = 0.02 and "Apoptosis" with an adjusted p-value = 1.025×10^{-2}. These pathways are associated with diseases that fall under the category of ear diseases on Kegg Disease' like "deafness" and "bilateral sudden sensorineural hearing loss". Additionally, the secondary targets are enriched in the PI3K-Akt signaling pathway (adjusted p-value = 4.28×10^{-5}); this pathway has been found to be an intrinsic protective mechanism of the inner ear. In particular, Chen et al. found that blockade of PI3K/Akt signaling pathways increases sensitivity to TTS noise-induced hearing loss [43].

In Fig. 5, we provide a graphical representation of the module under investigation. We highlighted the proteins involved in pathways of interest (here with a red outline). It is worth noting that they are not only direct neighbors of the targets. In particular, two of them (ARAF and VDAC1) are not first neighbors of any targets. Ignoring such proteins with distance from the targets greater than 1 would result in the module no longer being enriched in certain pathways. This is of particular interest because it shows how simply gathering all the first neighbors of the targets would bring to lose meaningful information. Moreover, given how few the possible targets are for this therapeutic function and how they are reused by many drugs, simply considering all the first neighbors will bring us to have the same proteins (or subsets of them) for each side effect, making the enrichment results undistinguishable and therefore useless.

3.3.6. Gastrointestinal issues

This side effect is common to 10 different drugs and represents another example of how the secondary targets help us have a clearer idea of why it can happen. Its targets appear not to be connected to the gastrointestinal apparatus or enriched in related pathways. However, the second result for the pathway enrichment of the secondary targets is the "retrograde endocannabinoid signaling pathway" (adjusted p-value $= 1 \times 10^{-4}$). This pathway has been previously linked with gastrointestinal issues: more specifically, some studies reveal an important (and at times surprising) role for the endocannabinoid system in the control of a variety of gastrointestinal functions, including motility, gut–brain mediated fat intake and hunger signaling, inflammation and gut permeability, and dynamic interactions with gut microbiota [44].

3.3.7. Thromboembolic diseases

Scientific and public attention to thromboembolism and hormonal contraception has had dramatic consequences, both good and bad. The spotlight on risk has helped to change norms regarding the public's right to know and assess dangers, but it also spiked pill scares, leading to increased unplanned pregnancy, birth and abortion rates [45]. This side effect looks to be caused mostly by drug targets, which are related to "Venous Thrombosis" (adjusted p-value = 4.05×10^{-3}), and of course to the "Estrogen signaling pathway" (adjusted p-value 9.043×10^{-6}), which has the power to increase the risk of both arterial and venous thrombosis [46].

4. Conclusion

In this study we proposed a computational approach for dealing with one of the main concerns in drug development and distribution: the occurrence of side effects. Indeed, predicting side effects and explaining

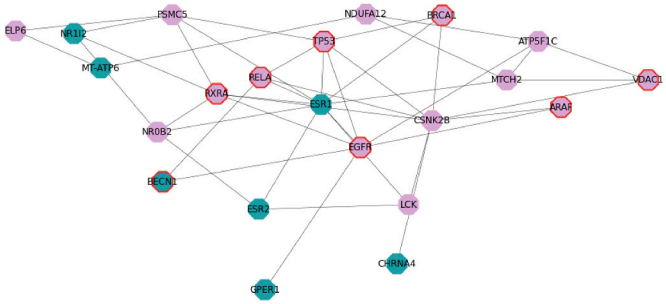


Fig. 5. The ear disorder side effect module. The module is composed of targets (in teal) and secondary targets (in lilac). The nodes with a red outline are the proteins enriched in pathways of interest for this side effect. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

what causes them can improve the patient experience and stop unsafe drugs from entering the market. Following previous studies which tried to understand side effects by joining information about drugs, targets, biological pathways and the human interactome, we carried out an in-depth case study about hormonal contraceptives and implemented a new method to investigate the molecular mechanisms of side effects. Differently from previous works, we focused on a specific category of drugs to obtain information that are therapy-specific and could easily be used in real life situations. Moreover, we considered exclusively side effects which are classified as common or very common, because they are the ones which have the biggest impact on the patient's life and their frequency assures that they cannot be caused by an individual response caused by genetic or phenotypic factors. To carry out this analysis we gathered and integrated data from many online databases and manually annotated side effects for the considered drugs (HCs) to then build and study the side effect modules. The enrichment analysis revealed that proteins composing the identified modules are associated with diseases/phenotypes and/or KEGG pathways related to the SEs. In particular, as expected, target proteins are mostly enriched in pathways related to the HCs therapeutic functions and in some cases, they also appear related to the side effects. Instead in other cases, targets do not explain SEs, while investigating their neighborhood (defined as the proteins that allow the targets' connection), we found SE-related pathways. The potential of the proposed method is that it can be applied to any category of drugs. Further developments are certainly needed to consider also other functional features of the module proteins that can explain the side effects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Sapienza University of Rome, "Progetto di Ateneo", grant number: RM12117A8B3C3033.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2023.101163.

References

[1] Hall KS, Castaño PM, Westhoff CL. The influence of oral contraceptive knowledge on oral contraceptive continuation among young women. J Womens Health Jul. 2014;23(7):596–601. https://doi.org/10.1089/jwh.2013.4574.

[2] Williams NM, Randolph M, Rajabi-Estarabadi A, Keri J, Tosti A. Hormonal contraceptives and dermatology. Am J Clin Dermatol Jan. 2021;22(1):69–80. https://doi.org/10.1007/s40257-020-00557-5.

[3] Tepper NK, Whiteman MK, Zapata LB, Marchbanks PA, Curtis KM. Safety of hormonal contraceptives among women with migraine: a systematic review. Contraception 2016;94(6):630–40. https://doi.org/10.1016/j.contraception.2016.04.016.

[4] Skovlund CW, Mørch LS, Kessing LV, Lidegaard Ø. Association of hormonal contraception with depression. JAMA Psychiatr 2016;73(11):1154–1162, Nov. https://doi.org/10.1001/jamapsychiatry.2016.2387.

[5] Shufelt CL, Bairey Merz CN. Contraceptive hormone use and cardiovascular disease. J Am Coll Cardiol Jan. 2009;53(3):221–31. https://doi.org/10.1016/j.jacc.2008.09.042.

[6] Cortés ME, Alfaro AA. The effects of hormonal contraceptives on glycemic regulation. Linacre Q Aug. 2014;81(3):209–18. https://doi.org/10.1179/2050854914Y.0000000023.

[7] of O. The commissioner, "the drug development process. 2020. *FDA*, Feb. 20, https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process. [Accessed 7 September 2022].

[8] Kuhn M, et al. Systematic identification of proteins that elicit drug side effects. Mol Syst Biol 2013;9:663. https://doi.org/10.1038/msb.2013.10.

[9] Lee S, Lee KH, Song M, Lee D. Building the process-drug-side effect network to discover the relationship between biological processes and side effects. BMC Bioinf Mar. 2011;12(Suppl 2):S2. https://doi.org/10.1186/1471-2105-12-S2-S2.

[10] Chen X, et al. Systematic analysis of the associations between adverse drug reactions and pathways. BioMed Res Int 2015;2015:670949. https://doi.org/10.1155/2015/670949.

[11] Guney E. Investigating side effect modules in the interactome and their use in drug adverse effect discovery. In: Complex networks VIII; 2017. p. 239–50.

[12] Chen X, et al. Network characteristic analysis of ADR-related proteins and identification of ADR-ADR associations. Sci Rep 2013;3:1744. https://doi.org/10.1038/srep01744.

[13] Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res 2016;44(D1):D1075–9. https://doi.org/10.1093/nar/gkv1075.

[14] Chen EY, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf Apr. 2013;14:128. https://doi.org/10.1186/1471-2105-14-128.

[15] Kuleshov MV, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44(W1):W90–7. https://doi.org/10.1093/nar/gkw377. 08.

[16] Xie Z, et al. Gene set knowledge discovery with Enrichr. Curr Protoc Mar. 2021;1(3):e90. https://doi.org/10.1002/cpz1.90.

[17] Wishart DS, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018;46(D1):D1074–82. https://doi.org/10.1093/nar/gkx1037. 04.

[18] Lynch T, Price A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. Am Fam Physician Aug. 2007;76(3):391–6.

[19] Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. Nat Commun Mar. 2019;10(1):1197. https://doi.org/10.1038/s41467-019-09186-x.

[20] Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med Mar. 2012;4(125):125ra31. https://doi.org/10.1126/scitranslmed.3003377.

[21] Drugs A-Z list - a on RxList," RxList. https://www.rxlist.com/drugs/alpha_a.htm. [Accessed 7 September 2022].

[22] Generic drugs. Search synonyms, analogues or drug substitutes. Price comparison.," *ndrugs*, https://www.ndrugs.com/. [Accessed 7 September 2022].

[23] Gahr M, Eller J, Connemann BJ, Schönfeldt-Lecuona C. Subjective reasons for non-reporting of adverse drug reactions in a sample of physicians in outpatient care. Pharmacopsychiatry Mar. 2016;49(2):57–61. https://doi.org/10.1055/s-0035-1569291.

[24] Gahr M, Connemann B, Zeiss R, Schönfeldt-Lecuona C, Dreyhaupt J, Lazik C. Reporting, handling, and subjective importance of adverse drug reactions among general practitioners: an exploratory cross-sectional survey. Expet Opin Drug Saf Aug. 2021;20(8):979–85. https://doi.org/10.1080/14740338.2021.1933430.

[25] http://www.laboratoryequipment.com/563770-Subjective-Side-Effect-Reporting-Clouds-Drug-Safety-in-Clinical-Trials/.

[26] Merckelbach H, Dandachi-FitzGerald B, van Helvoort D, Jelicic M, Otgaar H. When patients overreport symptoms: more than just malingering. Curr Dir Psychol Sci 2019;28(3):321–6. https://doi.org/10.1177/0963721419837681.

[27] C. for D. E and Research. The FDA's drug review process: ensuring drugs are safe and effective. FDA; Jun. 2020. Accessed: Sep. 07, 2022. [Online]. Available: https://www.fda.gov/drugs/information-consumers-and-patients-drugs/fdas-drug-review-process-ensuring-drugs-are-safe-and-effective.

[28] Wang R-S, Loscalzo J. Network-based disease module discovery by a novel seed connector algorithm with pathobiological implications. J Mol Biol Sep. 2018;430(18):2939–50. https://doi.org/10.1016/j.jmb.2018.05.016. Part A.

[29] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet Jan. 2011;12(1):56–68. https://doi.org/10.1038/nrg2918.

[30] Farina L. Network as a language for precision medicine. Ann Ist Super Sanita Dec. 2021;57(4):330–42. https://doi.org/10.4415/ANN_21_04_08.

[31] Paci P, Fiscon G, Conte F, Wang R-S, Farina L, Loscalzo J. Gene co-expression in the interactome: moving from correlation toward causation via an integrated approach to disease module discovery. NPJ Syst Biol Appl Jan. 2021;7(1):1–11. https://doi.org/10.1038/s41540-020-00168-0.

[32] Ghiassian SD, Menche J, Barabási A-L. A DIseAse MOdule detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLoS Comput Biol Apr. 2015;11(4):e1004120. https://doi.org/10.1371/journal.pcbi.1004120.

[33] Petti M, Bizzarri D, Verrienti A, Falcone R, Farina L. Connectivity significance for disease gene prioritization in an expanding universe. IEEE ACM Trans Comput Biol Bioinf 2020;17(6):2155–2161, Nov. https://doi.org/10.1109/TCBB.2019.2938512.

[34] Petti M, et al. MOSES: a new approach to integrate interactome topology and functional features for disease gene prediction. Genes 2021;12(no. 11):27. https://doi.org/10.3390/genes12111713.

[35] Allen J, Romay-Tallon R, Brymer KJ, Caruncho HJ, Kalynchuk LE. Mitochondria and mood: mitochondrial dysfunction as a key player in the manifestation of depression. Front Neurosci 2018;12:386. https://doi.org/10.3389/fnins.2018.00386.

[36] Catena-Dell'Osso M, Bellantuono C, Consoli G, Baroni S, Rotella F, Marazziti D. Inflammatory and neurodegenerative pathways in depression: a new avenue for antidepressant development? Curr Med Chem 2011;18(2):245–55. https://doi.org/10.2174/092986711794088353.

[37] Lutz B, Marsicano G, Maldonado R, Hillard CJ. The endocannabinoid system in guarding against fear, anxiety and stress. Nat Rev Neurosci 2015;16(12):705–18. https://doi.org/10.1038/nrn4036.

[38] Dehghan F, Haerian BS, Muniandy S, Yusof A, Dragoo JL, Salleh N. The effect of relaxin on the musculoskeletal system. Scand J Med Sci Sports Aug. 2014;24(4):e220–9. https://doi.org/10.1111/sms.12149.

[39] Gallo MF, Lopez LM, Grimes DA, Schulz KF, Helmerhorst FM. Combination contraceptives: effects on weight. Cochrane Database Syst Rev Sep. 2011;9:CD003987. https://doi.org/10.1002/14651858.CD003987.pub4.

[40] Lopez LM, et al. Progestin-only contraceptives: effects on weight. Cochrane Database Syst Rev Aug. 2016;8:CD008815. https://doi.org/10.1002/14651858.CD008815.pub4.

[41] Candler T, Kühnen P, Prentice AM, Silver M. Epigenetic regulation of POMC; implications for nutritional programming, obesity and metabolic disease. Front Neuroendocrinol 2019;54:100773. https://doi.org/10.1016/j.yfrne.2019.100773.

[42] Johansson LE, et al. Differential gene expression in adipose tissue from obese human subjects during weight loss and weight maintenance. Am J Clin Nutr Jul. 2012;96(1):196–207. https://doi.org/10.3945/ajcn.111.020578.

[43] Chen J, Yuan H, Talaska AE, Hill K, Sha S-H. Increased sensitivity to noise-induced hearing loss by blockade of endogenous PI3K/Akt signaling. J Assoc Res Otolaryngol JARO Jun. 2015;16(3):347–56. https://doi.org/10.1007/s10162-015-0508-x.

[44] DiPatrizio NV. Endocannabinoids in the gut. Cannabis Cannabinoid Res Feb. 2016;1(1):67–77. https://doi.org/10.1089/can.2016.0001.

[45] Lackie E, Fairchild A. The birth control pill, thromboembolic disease, science and the media: a historical review of the relationship. Contraception 2016;94(4):295–302. https://doi.org/10.1016/j.contraception.2016.06.009.

[46] Abou-Ismail MY, Citla Sridhar D, Nayak L. Estrogen and thrombosis: a bench to bedside review. Thromb Res 2020;192:40–51. https://doi.org/10.1016/j.thromres.2020.05.008.