INDI: A computational framework for inferring drug interactions and their associated recommendations

Assaf Gottlieb, Gideon Y. Stein, Yoram Oron, Eytan Ruppin, Roded Sharan

Corresponding author: Assaf Gottlieb, Tel-Aviv University

Review timeline:

| Event                  | Date       |
|------------------------|------------|
| Submission date        | 15 February 2012 |
| Editorial Decision     | 21 March 2012   |
| Revision received      | 03 May 2012   |
| Editorial Decision     | 21 May 2012   |
| Revision received      | 01 June 2012   |
| Accepted               | 04 June 2012   |

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 21 March 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees who agreed to evaluate your manuscript. As you will see from the reports below, the referees find the topic of your study of potential interest. They raise, however, substantial concerns on your work, which, I am afraid to say, preclude its publication in its present form.

Broadly, the reviewers appreciated the goals of this work; however, all three reviewers clearly felt that the drug interaction predictions required more rigorous, independent testing and validation. The last reviewer felt that this would require comparisons to additional independent datasets, and the second reviewer specifically suggests using the clinical data from which the prescribed DDI prevalences were estimated. The first reviewer was far more critical, but his/her concerns largely emphasize the need for additional independent clinically-relevant validation of these predictions. The editor would like to stress these concerns, and the apparent agreement between the reviewers.

The editor acknowledges that the first reviewer's report is rather bluntly worded, but I think s/he raises some valuable points that require serious consideration. You emphasize that these DDI predictions can be used as clinically-relevant "recommendations", and you write that the web tool, "may aid physicians and researchers to exploit INDI's predictions in the clinical practice." Since these predictions do not account for dosages, and may not distinguish therapeutically beneficial drug synergies from adverse combination effects, the actual clinical action merited for each combination, even when a DDI is likely, will often be unclear. Substantial caution is merited in this regard, since, as the first reviewer points out, a direct reading of these predictions by physicians could lead to the avoidance of potentially effective drug combinations in clinical practice.

During the cross-commenting period, a reviewer also noted the recent publication by Tatonetti et al
(2012, Sci Transl Med), as an example of a work that includes the kind of validation with clinical data that they felt would greatly improve this manuscript.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favorable.

*PLEASE NOTE* As part of the EMBO Publications transparent editorial process initiative (see http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology now publishes online a Review Process File with each accepted manuscript. Please be aware that in the event of acceptance, your cover letter/point-by-point document will be included as part of this file, which will be available to the scientific community. Authors may opt out of the transparent process at any stage prior to publication (contact us at msb@embo.org). More information about this initiative is available in our Instructions to Authors.

Sincerely,

Editor - Molecular Systems Biology
msb@embo.org

---------------------------------------------------------------------------

Reviewer reports

Reviewer #1 (Remarks to the Author):

This manuscript provides an extension of previously published work in the same journal. It provides vast amounts of pre-processed, putative drug interaction alerts that can not be easily verified or ascertained in clinically relevant conditions.

The supplementary material includes 1300 pages of tables related to potential drug interactions, most of which indicating "monitor", or "avoid", or "adjust dosage". Since no dosage appears to be factored into the calculation, it is not clear to this reviewer if dosage should be adjusted higher, or lower (presumably lower).

Surprisingly, of the 84 drug metabolizing enzymes, only six cytochrome P450 isozymes these have been deemed relevant for prediction in spite of clear evidence pointing to the interplay with efflux transporters, food effects (which influence AUC drug levels) and other enzymes such as aldehyde oxidase 1, as potential clinically relevant culprits for drug interactions.

No discrimination is given between substrates and inhibitors of the P450s; in particular 3A4 and 2D6 can handle different substrates, with different Km values, at the same time. Therefore, just because two drugs are metabolized by the same enzyme, does not mean an interaction alert should be issued.

Some random examples are monitored below:

really bad predictions
1. Vincristine & doxorubicin have been combined, and showed more effective compared to single drug.

2. Fluorouracil and methotrexate: "Adjuvant combination chemotherapy that includes methotrexate, cyclophosphamide and fluorouracil has been used most extensively and is considered a regimen of choice" (followed by 16 references) according to the American Society of Health-System Pharmacists. This is contrary to what the authors recommend.

puzzling predictions
3. Indomethacin is metabolized by CYP2C19, an enzyme not influenced by caffeine; so why is their combination to be "monitored"?! people drink coffee while taking indomethacin on a daily basis...
4. Rifampin is metabolized by more than 6 P450s, whereas mercaptopurine is metabolized by xanthine dehydrogenase/oxidase; in the unlikely event where one would combine a tuberculostatic with an immunosuppressant / antineoplastic, it is unclear why this drug combination could cause additional alerts.

5. Gefitinib and irinotecan have been combined by at least one group, with no clear therapeutic benefit; however the combination did not cause serious drug interactions.

6. Desonide is a topical cream with no evidence of systemic absorption in usual doses (as searched on 3 separate databases). Why is this even flagged for interactions?!

Furthermore, of major drug interaction alerts, QT prolongators, NSAIDs & COXibs, and coumarins are among the most frequently encountered; yet no attention is paid to those generally accepted warning flags.

Somewhat disturbing, no thought is given to the medical and ethical consequences of this "prediction fest". What would happen if these predictions, not rooted in evidence, fall in the hands of lawyers who use them in trials, placing the burden of proof on physicians, pharmacists and hospitals?!

Reviewer #2 (Remarks to the Author):

The article by Gottlieb et al. describes an algorithm (INDI), based on known drug-drug interactions (DDI) and drug similarities, which predicts new drug-drug interactions. The reported quality controls strongly suggest that the algorithm is successful for inferring DDIs, hence it can be useful for better treatment. I am enthusiastic about the article, however I have a set of criticisms. I would be very happy to reevaluate the article after these criticisms are considered.

1. The DDI data that is used for training the algorithm is taken from drugbank and drugs.com. However, the authors do not provide the DDI data as part of this article. I believe this should be appended since the reader should be in a position to repeat the analysis if desired.

2. There is a lot of data integration, cross checks and predictions within the article. However, there is no figure that visually summarizes the flow of the article and reflects the number of data points or predictions in each step. I believe such a figure would greatly increase the readability. For example, I would like to see the # of DDIs coming from two data types, the # of DDIs found in both, the # of DDIs predicted for each type, etc.

3. The authors start their analysis by partitioning the known DDIs into 3 types using the DDI's relationship with CYP. As a reader, I would first want to be shown that the INDI algorithm works - regardless of the CYP, since the CYP relation is not used in the prediction scheme. If it does, then the next step would be to check does it work better when different DDI types are considered; if this next step is not a significant improvement, the partitioning of the DDI data might be done later when CYPs are predicted for CRDs.

4. A major concern is that, although the paper predicts DDIs, it never formally defines what a DDI is. Making training data available as requested in point 1 will partly remedy this situation, but a proper definition with examples will increase the relevance of the article.

5. The second paragraph of the results, which relates to the trends in DDI properties, I believe would be more appropriately placed in Supplementary Information (SI) section 1, where a large number of similar observations are done. I find these observations (both in results and SI) very interesting, however the current version of this paragraph is not very convincing (p values around 0.03 are not significant when multiple comparisons are done) and the findings in the paragraph do not really
seem to warrant a figure.

6. INDI website gives DDI predictions, but doesn't give real DDIs listed in drugbank or drugs.com. I think having a separate section where known real DDIs are given could be a good idea.

7. It is stated: "The first [assessment of predictions] is based on the assumption that NCRDs tend to occur among drugs affecting the same tissues." Why? I would expect the opposite, because NCRDs presumably affect different proteins.

8. Figure 3 is very cryptic since the key for the abbreviations is not given either on the figure or legend. However, I am very interested in the tendency of frequent DDIs across drug classes. The figure and the related text could use an update.

9. I would very much prefer if the p values are given with an = sign instead of a < sign.

10. The correspondence between ADR database and predicted DDIs is very striking. However, the authors do not provide any evidence that these two data sets are independent. If drugbank and ADR database use each other's data, then it would not be surprising that they predict each other.

11. An extremely interesting part of the paper is when it presents the prevalence of the usage of drug combinations with DDIs, in real patients in a clinical setting. I am very interested in the following question and I really hope that it can be answered using the medical record data that they used for this estimation: What happened to the patients that used drug combinations with DDIs? Did they have different remissions than others?

12. Authors' claim of novelty in unbiased prediction of DDIs is not warranted. There is a growing literature in this field, including the following articles that appeared recently in MSB: Nelander et al 2008, Jansen et al 2009.

Reviewer #3 (Remarks to the Author):

In this paper Gottlieb and collaborators present a method to predict drug-drug interactions. The method uses an algorithm employed recently by the same authors to predict drug indications. They infer both pharmacokinetic and pharmacodynamics interactions, the severity level of the interaction and the CYP member family responsible for the interactions for the pharmacokinetic interactions. In addition, the authors developed a web tool to search for predicted interactions with information about the predicted recommendation and CYP metabolizing enzymes involved for pharmacokinetic interactions. The paper is well written and tackles an important issue in drug treatment. The method performs very well in cross-validation, however, the performance of the method in independent datasets and to what extend the predictions are not trivial predictions need further clarification.

Comments:

1. To validate the performance of the method in an independent dataset, the authors trained the method with the DrugBank dataset and validate it with interactions from www.drugs.com. It is shown that the fraction of predicted interactions in www.drugs.com (10% for CRD and 25% for NCRD) for the best F1-measure is significantly enriched in www.drugs.com dataset. Given the size of www.drugs.com set, which is 7-fold bigger than DrugBank set, one would expect to see more than 10% and 25% prediction overlap. To have a better idea of the performance of the method in the www.drugs.com dataset it would be interesting to see the ROC plot produced using the www.drugs.com interactions as benchmark set (restricting this set to the drugs present in DrugBank) and compare it to the ROC plot from cross-validation.

2. In the abstract it is stated "18% of the interactions of hospitalized patients are predicted to receive severely interactions drugs". According to the text, 18% is the percentage of patients with known or
predicted severe DDI but not with percentage of patients predicted to receive severe interacting drugs (1.3% CRD, 4.4% NCRD) mentioned later on in the manuscript. The authors should correct the abstract.

3-Page 12, paragraph 1: it is unclear to me whether the percentage of patients predicted to receive severe interacting drugs (1.3% CRD, 4.4% NCRD) is significantly higher than a random expectation. If it is not significant, the statement of the work demonstrating the prevalence of the predictions in regularly (page 4, paragraph 2) taken medications reported by hospitalized patients should be downgraded.

4-Regarding the DDI predicted on FDA adverse reporting data, in the discussion it is said, "our predictions significantly cover the majority of the adverse pairs of drugs reports". According to the text, only 4% (CRD) and 8% (NCRD) of the drug pairs are novel predictions, therefore, "the majority of the predictions" is an overestimation.

5-We have searched in www.drugs.com the ten best predicted CRDs. Three of them are already in www.drugs.com interaction checker (Carbamazepine-Trimipramine, Praziquantel-amobarbital and Triamcinolone-Guanfacine) and another three are drug pairs very similar to other DDI in www.drugs.com (Levofoxacin-Oxtirphyline similar to ofloxacin oxtriphylline, Triamcinolone-Guanfacine similar to triamcinolone guanfacine, Desonide Guanfacine similar to guanfacine budesonide). The authors should make sure the predictions are not part of the training set. In addition, in order to evaluate the novelty of the predictions it would be interesting to know how many predictions are not just trivial predictions, that is, drug pairs with high structural similarity to known DDI pairs.

6-They authors used CDK fingerprints to calculate chemical-based similarities and ECP4 and Daylight for ligand based similarities. In the manuscript it is not written which of the fingerprints is employed to construct the non-redundant set of chemicals. In order to assess the performance of the method with this non-redundant set it would be desired to use the same fingerprints in all the steps of the method as it might be that for example, a non-redundant set constructed with CDK fingerprints retains some structural similarity that will be captured by ECP4 or Daylight fingerprints in the ligand-based similarities.

7-Regarding the method: The description of the classification features and the query process is not easy to understand. For instance, it appears that when comparing the query drug pair to a drug pair with known interaction, one query drug is compared to only one drug in the drug pair with known interaction (by means of the seven different measures), so, d1 only with d1' and d2 with d2'. This implies that (d1',d2') is rather an ordered tuple than a set, and that there is also a corresponding drug pair (d2',d1').

8-The authors report to use a semantic similarity approach for ATC and GO terms according to Resnik. It would be interesting to know which particular approach is used.

9-From the viewpoint of bias it should also be considered if an approach based in the information content (probability p(x) at (4) annotation-based) is appropriate for the ATC-based annotation approach, as I would expect a bias towards those drug classes which contain less drugs, and therefore to the diseases for which there are currently not many drugs.
The Editor:

Broadly, the reviewers appreciated the goals of this work; however, all three reviewers clearly felt that the drug interaction predictions required more rigorous, independent testing and validation…

Following the reviewers’ comments we added additional validations as detailed below. In particular, using EMR data we now show that patients receiving medications with both known and predicted severe interactions tend to be hospitalized more frequently than patients who do not (2.2 times on average over a course of a year compared to only 1.6, p<4e-11, page 13). Importantly, there is no difference in age, gender or the hospitalization primary discharge diagnosis between the different groups of patients (i.e. receiving or not receiving drugs with severe interactions). In addition, we show (page 10) that our predictions overlap 40% of the independently extracted DDIs of Tatonetti et al., 2012 (p<9e-54) as well as his set of EMR-corroborated DDIs (p<0.035).

The editor acknowledges that the first reviewer’s report is rather bluntly worded, but I think s/he raises some valuable points that require serious consideration. You emphasize that these DDI predictions can be used as clinically-relevant “recommendations”, and you write that the web tool, "may aid physicians and researchers to exploit INDI’s predictions in the clinical practice." Since these predictions do not account for dosages, and may not distinguish therapeutically beneficial drug synergies from adverse combination effects, the actual clinical action merited for each combination, even when a DDI is likely, will often be unclear. Substantial caution is merited in this regard, since, as the first reviewer points out, a direct reading of these predictions by physicians could lead to the avoidance of potentially effective drug combinations in clinical practice.

Following this comment we now acknowledge more explicitly the limitations of our predictions and discuss in more detail and caution their potential exploration (page 17).

During the cross-commenting period, a reviewer also noted the recent publication by Tatonetti et al (2012, Sci Transl Med), as an example of a work that includes the kind of validation with clinical data that they felt would greatly improve this manuscript

We thank the reviewer for pointing us to this valuable source and include now a clinical analysis as well as an additional comparison to Tatonetti et al. EMR-corroborated set (page 10).

Reviewer 1:

The supplementary material includes 1300 pages of tables related to potential drug interactions, most of which indicating “monitor”, or “avoid”, or “adjust dosage”. Since no dosage appears to be factored into the calculation, it is not clear to this reviewer if dosage should be adjusted higher, or lower (presumably lower).

Following the reviewer’s comment, we further partitioned the dosage adjustment recommendations into four categories: increase, decrease, limit of dosage and adjustment of dosage interval. We next learned a classification rule for all these sub-categories of “adjust dosage” recommendations, providing additional evidence on the type of dosage adjustments needed (AUC>0.84, Tables 2 and S5 and page 11). We note that most of the known DDIs do not specify the exact amount of dosage adjustment, allowing the prediction of the direction only (increase, decrease) but not its magnitude.

Surprisingly, of the 84 drug metabolizing enzymes, only six cytochrome P450 isozymes these have been deemed relevant for prediction in spite of clear evidence pointing to the interplay with efflux transporters, food effects (which influence AUC drug levels) and other enzymes such as aldehyde oxidase 1, as potential clinically relevant culprits for drug interactions.

Our method is generic and could potentially predict other metabolizing enzymes. However, relating to the different CYP isozymes, 90.5% of our CRDs are listed in Drugs.com to be associated with one of the seven CYP isozymes, 9% additional ones are listed in Drugs.com as stemming from a CYP enzyme without specifying the exact isozyme involved, and only 0.5% are metabolized by rare isozymes that were not included in our prediction set. Other, non-CYP metabolizing enzymes were
excluded from the analysis due to the low number of such interactions, prohibiting proper prediction, as noted in the manuscript (page 19). Specifically, aldehyde oxidase mentioned by the reviewer is associated with only one interaction in our data (cimetidine and zaleplon).

No discrimination is given between substrates and inhibitors of the P450s; in particular 3A4 and 2D6 can handle different substrates, with different Km values, at the same time. Therefore, just because two drugs are metabolized by the same enzyme, does not mean an interaction alert should be issued.

As specified in the manuscript (page 5), for a DDI to be considered CYP-related we required that the interaction description specifically reports a CYP cause for it (in addition to the requirement that the two drugs are metabolized by the same CYP).

Some random examples are monitored below:

We agree with the reviewer that the benefit of a drug combination may outweigh the adverse effect of a possible interaction between them. And yet, in the same token, it is important to note the potential interaction and its type and severity, so that the physician can weigh all the pertaining information and make the best informed decision on a case by case basis. We now explicitly relate to the limitations of our approach in the discussion (page 17). In the following we address the specific examples provided by the reviewer.

really bad predictions
1. Vincristine & doxorubicin have been combined, and showed more effective compared to single drug.

The recommendation of this prediction is "monitor". This prediction is indeed reported in Drugs.com to be a true interaction, albeit with a minor effect (and thus was not included in our training set), suggesting that "caution and close observation are advised" (http://www.drugs.com/drug-interactions/adriamycin-with-vincristine-938-503-2301-0.html)

2. Fluorouracil and methotrexate: "Adjuvant combination chemotherapy that includes methotrexate, cyclophosphamide and fluorouracil has been used most extensively and is considered a regimen of choice" (followed by 16 references) according to the American Society of Health-System Pharmacists. This is contrary to what the authors recommend.

We do not dispute the benefits of this combination. However, this combination was indeed found to be mutually antagonistic under some circumstances and maintaining a 24 hour period between administrations of the two drugs was found to be preferable, as was specified in the manuscript (page 15).

puzzling predictions
3. Indomethacin is metabolized by CYP2C19, an enzyme not influenced by caffeine; so why is their combination to be "monitored"?! people drink coffee while taking indomethacin on a daily basis...

We predicted this interaction to be pharmacodynamic, which is unrelated to CYP isozymes. We note that there actually have been reports of interactions between Indomethacin and caffeine (e.g. “Interaction of ibuprofen and indomethacin with caffeine”, RM Khalil, Die Pharmazie, 1996).

4. Rifampin is metabolized by more than 6 P450s, whereas mercaptopurine is metabolized by xanthine dehydrogenase/oxidase; in the unlikely event where one would combine a tuberculostatic with an immunosuppressant / antineoplastic, it is unclear why this drug combination could cause additional alerts.

We predicted this interaction to be pharmacodynamic, which is unrelated to CYP isozymes. While we could not find literature evidence for such an interaction, possibly due to the rarity of the combination, we note that closely related (structurally and functionally) antimetabolite antineoplastic agents like Thioguanine are reported in Drugs.com to interact with Rifampin by inducing hepatotoxicity.
5. Gefitinib and irinotecan have been combined by at least one group, with no clear therapeutic benefit; however the combination did not cause serious drug interactions. 

As stated in Drugs.com regarding the interaction of Gefitinib and a CYP 3A4 inhibitor Atazanavir: “Coadministration with inhibitors of CYP450 3A4 may increase the plasma concentrations of gefitinib, which is primarily metabolized by the isozyme”. According to DrugBank and SuperCYP, Irinotecan is also a CYP3A4 inhibitor potentially producing the same effect as Atazanavir. We also note that it was previously reported that Gefitinib increases the bioavailability of orally administered Irinotecan, however the suspected mechanism of CYP3A4 could not be affirmed (Ken- ichi Fujita et al. Drug Metabolism and Disposition, 2005).

6. Desonide is a topical cream with no evidence of systemic absorption in usual doses (as searched on 3 separate databases). Why is this even flagged for interactions?!

Addressing the reviewer's comment, we removed interactions involving topical drugs (page 16) - thanks.

Furthermore, of major drug interaction alerts, QT prolongators, NSAIDs & COXibs, and coumarins are among the most frequently encountered; yet no attention is paid to those generally accepted warning flags.

One of the features of our method is that promiscuous drug classes (i.e., drug classes known to be involved in multiple interactions) tend to have higher number of predicted interactions. This property is manifested by the high correlation between the number of known and predicted drug interactors per drug (Supplementary material, page 2). It is thus unnecessary to pre-define promiscuous drug classes (e.g. QT prolongators, NSAIDS, etc). Obviously, it was of importance to keep the prediction method unbiased and avoid including such information beforehand.

Somewhat disturbing, no thought is given to the medical and ethical consequences of this "prediction fest". What would happen if these predictions, not rooted in evidence, fall in the hands of lawyers who use them in trials, placing the burden of proof on physicians, pharmacists and hospitals?!

We thank the reviewer for this thoughtful comment. We now describe our predictions with more caution and detail, explaining their caveats and limitations, both in the manuscript (page 19) and on our web site. We further suggest (in the discussion) that our predictions may be beneficial in three areas: (i) drug development, especially in post-marketing surveillance, aiding in verification of hazardous interactions; (ii) large scale clinical trial design, addressing and assessing potentially hazardous drug combinations; and (iii) driving and directing in-vitro validation of potentially hazardous interactions for efficiency and cost reduction of large scale biological experiments.

Reviewer 2:

1. The DDI data that is used for training the algorithm is taken from drugbank and drugs.com. However, the authors do not provide the DDI data as part of this article. I believe this should be appended since the reader should be in a position to repeat the analysis if desired.

The DrugBank data is freely available for download at www.drugbank.ca/downloads (which we also note in the manuscript, page 19). However, according to the terms of use of www.drugs.com (http://www.drugs.com/support/terms.html), no material from their website may be “republished, posted, distributed in any way, or incorporated into any other website”.

2. There is a lot of data integration, cross checks and predictions within the article. However, there is no figure that visually summarizes the flow of the article and reflects the number of data points or predictions in each step. I believe such a figure would greatly increase the readability. For example, I would like to see the # of DDIs coming from two data types, the # of DDIs found in both, the # of DDIs predicted for each type, etc.

We thank the reviewer for this wonderful suggestion and added the suggested overview figure (Figure 1).
3. The authors start their analysis by partitioning the known DDIs into 3 types using the DDI's relationship with CYP. As a reader, I would first want to be shown that the INDI algorithm works - regardless of the CYP, since the CYP relation is not used in the prediction scheme. If it does, then the next step would be to check does it work better when different DDI types are considered; if this next step is not a significant improvement, the partitioning of the DDI data might be done later when CYPs are predicted for CRDs.

We now show that the prediction of known DDIs yields high performance in cross validation scheme ($AUC=0.97\pm4e^{-4}$, page 8). However, the type of interaction (pharmacokinetic or pharmacodynamic) constitutes important information that is ignored in such a prediction scheme. This requires an additional inference layer – that of the type of interaction, introducing unnecessary complexity. Since the performance on each interaction type is comparable to the performance on all the interactions, we preferred to retain this type of information in exchange for a negligible performance reduction.

4. A major concern is that, although the paper predicts DDIs, it never formally defines what a DDI is. Making training data available as requested in point 1 will partly remedy this situation, but a proper definition with examples will increase the relevance of the article.

Relying on the definition found in Crowther et al., 1997, a DDI is defined as any drug effect greater/less than expected in the presence of another drug. We added this definition to the introduction (page 2) – thanks.

5. The second paragraph of the results, which relates to the trends in DDI properties, I believe would be more appropriately placed in Supplementary Information (SI) section 1, where a large number of similar observations are done. I find these observations (both in results and SI) very interesting, however the current version of this paragraph is not very convincing (p values around 0.03 are not significant when multiple comparisons are done) and the findings in the paragraph do not really seem to warrant a figure.

We moved part of this section and the corresponding figure to the supplements (supplementary material, page 3 and Figure S3). The p-values are for the Spearman ranked correlations and do not involve multiple comparisons.

6. INDI website gives DDI predictions, but doesn't give real DDIs listed in drugbank or drugs.com. I think having a separate section where known real DDIs are given could be a good idea.

We cannot add these data to our website, as it does not comply with the Drugs.com terms of use (see answer to comment #1 above). However, clicking the embedded links in the drug names retrieves the relevant DrugBank record, including all interactions associated with the drug in the DrugBank database.

7. It is stated: "The first [assessment of predictions] is based on the assumption that NCRDs tend to occur among drugs affecting the same tissues." Why? I would expect the opposite, because NCRDs presumably affect different proteins.

The basic assumption is that the interaction occurs since the drugs affect the same tissue, regardless of whether the same or different proteins are involved (e.g. both drugs affect heart tissue in QT-prolongation, hepatotoxicity in liver tissue etc.). This assumption was first validated using the known interactions (Supplementary Material, page 4).

8. Figure 3 is very cryptic since the key for the abbreviations is not given either on the figure or legend. However, I am very interested in the tendency of frequent DDIs across drug classes. The figure and the related text could use an update.

The abbreviations are standard 3rd level ATC classes. We added the top level (anatomical) classification to the legend.

9. I would very much prefer if the p values are given with an = sign instead of a < sign.
10. The correspondence between ADR database and predicted DDIs is very striking. However, the authors do not provide any evidence that these two data sets are independent. If drugbank and ADR database use each other's data, then it would not be surprising that they predict each other.

DrugBank is independent from the ADR reports as DrugBank collects interactions from literature curated databases whereas FDA ADR reports are individual patient adverse events reported by physicians without clear association between the adverse event and the drugs taken by the patient at the time (see also Tatonetti et al., 2012, see the manuscript's reference list for the complete reference).

11. An extremely interesting part of the paper is when it presents the prevalence of the usage of drug combinations with DDIs, in real patients in a clinical setting. I am very interested in the following question and I really hope that it can be answered using the medical record data that they used for this estimation: What happened to the patients that used drug combinations with DDIs? Did they have different remissions than others?

We thank the reviewer for this interesting suggestion. Following it, we verified that the patients having either known or predicted severe interactions were hospitalized more frequently than those patients who had none (page 13). In detail, patients having either known or predicted severe interactions were hospitalized 2±1.6 and 2.2±1.7 times on average during a course of a year, while patients who had none were hospitalized only 1.6±1.3 times on average (Wilcoxon ranked sum test p=4e⁻¹¹, and p=2e⁻¹⁸, respectively).

12. Authors' claim of novelty in unbiased prediction of DDIs is not warranted. There is a growing literature in this field, including the following articles that appeared recently in MSB: Nelander et al 2008, Jansen et al 2009.

We are unaware of other large scale in-silico DDI prediction methods. Specifically, Nelander et al. 2008 and Jansen et al. 2009 require in-vitro measurements of molecular profiles and are currently validated only at small scale. We now added these references and discuss their limitations in the introduction (page 3).

Reviewer #3 (Remarks to the Author):

1-To validate the performance of the method in an independent dataset, the authors trained the method with the DrugBank dataset and validate it with interactions from www.drugs.com. It is shown that the fraction of predicted interactions in www.drugs.com (10% for CRD and 25% for NCRD) for the best F1-measure is significantly enriched in www.drugs.com dataset. Given the size of www.drugs.com set, which is 7-fold bigger than DrugBank set, one would expect to see more than 10% and 25% prediction overlap. To have a better idea of the performance of the method in the www.drugs.com dataset it would be interesting to see the ROC plot produced using the www.drugs.com interactions as benchmark set (restricting this set to the drugs present in DrugBank) and compare it to the ROC plot from cross-validation.

Please note that the true coverage of Drugs.com known interactions are 25% and 41% for CRD and NCRD predictions, respectively, including known drugs.com PCRDs (supplementary information, page 4). We added the suggested validation, obtaining AUC=0.92 and 0.95 for CRDs and NCRDs, respectively on the cross validation (Table S4) and p<4e⁻⁴ when validating the DrugBank known interactions against predictions made on the Drugs.com set (page 9).

2- In the abstract it is stated "18% of the interactions of hospitalized patients are predicted to receive severely interacting drugs". According to the text, 18% is the percentage of patients with known or predicted severe DDI but not with percentage of patients predicted to receive severe interacting drugs (1.3% CRD, 4.4% NCRD) mentioned later on in the manuscript. The authors should correct the abstract.
We corrected the abstract accordingly - thanks.

3-Page 12, paragraph 1: it is unclear to me whether the percentage of patients predicted to receive severe interacting drugs (1.3% CRD, 4.4% NCRD) is significantly higher than a random expectation. If it is not significant, the statement of the work demonstrating the prevalence of the predictions in regularly (page 4, paragraph 2) taken medications reported by hospitalized patients should be downgraded.

The percentage of predicted severe interactions is indeed not higher than the random expectation, which is generally expected as the prescribing physicians have no knowledge of these potential interactions. We do expect, however, the frequency of known severe interactions, of which physicians are aware, to be slim. We now note that these percentages follow the random expectation (page 13). We note that in this revision the numbers have slightly changed after correcting for multiple hospitalizations of the same patient.

4-Regarding the DDI predicted on FDA adverse reporting data, in the discussion it is said, "our predictions significantly cover the majority of the adverse pairs of drugs reports". According to the text, only 4% (CRD) and 8% (NCRD) of the drug pairs are novel predictions, therefore, "the majority of the predictions" is an overestimation.

We meant that both known and predicted DDIs cover the majority of the adverse pairs of drugs reports. We corrected this sentence (page 17).

5-We have searched in www.drugs.com the ten best predicted CRDs. Three of them are already in www.drugs.com interaction checker (Carbamazepine-Trimipramine, Praziquantel-amobarbital and Triamcinolone-Guanfacine) and another three are drug pairs very similar to other DDI in www.drugs.com (Levofloxacin-Oxtriphylline similar to ofloxacin ↔ oxtriphylline, Triamcinolone-Guanfacine similar to triamcinolone ↔ guanfacine, Desonide Guanfacine similar to guanfacine ↔ budesonide). The authors should make sure the predictions are not part of the training set. In addition, in order to evaluate the novelty of the predictions it would be interesting to know how many predictions are not just trivial predictions, that is, drug pairs with high structural similarity to known DDI pairs.

No prediction is part of the training set. However, a significant portion of the predictions intersects PCRDs - a set of known DDIs that was left for validation (12% of predicted CRDs, p<5e^-313 and 28% of predicted NCRDs, p=0, supplementary material, page 4).

In order to evaluate the effect of structure similarity on the prediction performance, we indeed performed (already in the original manuscript) a cross validation on a non-redundant set, removing structurally similar drugs or drugs sharing the same targets, showing that the difference in AUC was slim (<0.03, seeTable 1).

6-They authors used CDK fingerprints to calculate chemical-based similarities and ECP4 and Daylight for ligand based similarities. In the manuscript it is not written which of the fingerprints is employed to construct the non-redundant set of chemicals. In order to assess the performance of the method with this non-redundant set it would be desired to use the same fingerprints in all the steps of the method as it might be that for example, a non-redundant set constructed with CDK fingerprints retains some structural similarity that will be captured by ECP4 or Daylight fingerprints in the ligand-based similarities.
The non-redundant set of chemicals was constructed based on the CDK fingerprints, which are used for calculation of the commonly used Tanimoto 2D coefficients (page 20). The CDK fingerprints are very similar to the commercial Daylight fingerprints (see “Steinbeck et al.”, J. Chem. Inf. Comput. Sci., 2003). While using Tanimoto coefficients as low as 0.5 is generally sufficient for producing a non-redundant set that does not include trivially similar drugs (e.g. Mason et al., J. Med. Chem., 1999), we repeated the cross validation on the non-redundant set excluding the ligand similarity to ensure there is no bias and showed that while the CRD results were similar, the NCRDs results were even better without the ligand similarity measure. The AUC difference were smaller than 0.01 (page 9 and Figure S1).

We note that the computation of the ECPF-4 and Daylight fingerprints for construction of the ligand based similarities was done by the Similarity Ensemble Approach web tool, prohibiting the usage of the CDK fingerprints for that purpose.

7-Regarding the method: The description of the classification features and the query process is not easy to understand. For instance, it appears that when comparing the query drug pair to a drug pair with known interaction, one query drug is compared to only one drug in the drug pair with known interaction (by means of the seven different measures), so, d1 only with d1’ and d2 with d2’. This implies that (d1’,d2’) is rather an ordered tuple than a set, and that there is also a corresponding drug pair (d2’,d1’).

We compare also d1 with d2’ and d2 with d1’. We now clarify this in the manuscript (page 22).

8-The authors report to use a semantic similarity approach for ATC and GO terms according to Resnik. It would be interesting to know which particular approach is used.

We used the approach suggested by Resnik et al. in the paper “Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language”, whereby the similarity of two drugs is the maximum over all the probability of their common ancestors c of \(-\log(p(c))\). We clarify this and provide the reference on page 21.

9-From the viewpoint of bias it should also be considered if an approach based in the information content (probability p(x) at (4) annotation-based) is appropriate for the ATC-based annotation approach, as I would expect a bias towards those drug classes which contain less drugs, and therefore to the diseases for which there are currently not many drugs.

As we show in the manuscript (Suppl. Table S4), removing the ATC similarity does not lead to significant performance degradation (AUC change < 0.02), suggesting that the ATC similarity measure does not substantially bias the results.

2nd Editorial Decision 21 May 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who agreed to evaluate this revised study. As you will see, the referees felt that the revisions made had improved this work. The last reviewer, however, has an important remaining concern that we feel is sufficient to preclude publication of this work, at present.

Specifically, the last reviewer felt it would be essential to clarify to what degree known DDIs contribute to the subsequent validating analyses, and to specifically test whether the novel predictions of this computational method, excluding all known DDIs, still show independent evidence of adverse interactions. While it appears that this issue can likely be addressed with some additional clarification and reanalysis, it does impact on the direct support for the value of this method, so we reserve the right to consult again with reviewers if needed, and, of course, we cannot guarantee a positive outcome at this point.

In addition, the second reviewer is still somewhat concerned about the ability of researchers to
replicate this analysis, since the full DDI training set is not included with this work. Despite the public availability of DrugBank and Drugs.com, I would assume that these resources will continue to evolve with time, which may make it hard for others to extract exactly the same training set you used here. Obviously, it would be ideal if you can distribute the full training set, and I would encourage you to contact DrugBank and Drugs.com to see if an exception can be granted. Nonetheless, I understand that this may not be possible, and in lieu of full release I would encourage you to describe in as much detail as possible the process by which data was extracted from these databases and partitioned for training and validation. The date on which the datasets were downloaded should also be stated.

When preparing your revised work, we ask that you also address the following format and content issues:

1. In general, we prefer to avoid titles with colons or abbreviations whenever possible, and discourage the use of "novel" or "new" in titles (since all works should be presenting novel findings or methods). As such, we encourage you to consider an alternate title when preparing this final revision, the simplest probably being, "A computational framework for inferring drug interactions and their associated recommendations".

2. Please provide the figure images in a format that preserves the lines and text as vectors -- ideally a direct EPS or PDF output from a program like Illustrator or Inkscape. as given on the submission website.

Thank you for submitting this paper to Molecular Systems Biology.

Sincerely,

Editor - Molecular Systems Biology
msb@embo.org

Referee reports

Reviewer #2 (Remarks to the Author):

The authors have satisfactorily addressed all my concerns I indicated in my review except the first one: the availability of drug interaction data sets used for the development of their drug interaction prediction tool. Authors state that first can be downloaded online. However, the second one is not freely available, so it can't be redistributed. I personally feel this is not a satisfactory response. Perhaps the authors can describe in detail how they obtained this data so readers may also reproduce the study if they wished to. I leave it to the editor's prudence to decide if my concern precludes publication. I believe that the article has significantly improved after the reviews. I am particularly moved by the verification of the drug interaction prediction tool, using medical records. I recommend its publication.

Reviewer #3 (Remarks to the Author):

Although the authors have clarified in this revision round some of my concerns, it is still unclear to me whether the DDIs predictions are meaningful or not. I am now confused about the meaning of "predictions" in this article. In the response to our second point, the authors acknowledge that a "significant portion of predictions intersect PCRDs - a set of known DDIs left for validation". Thus, this means that a high number of predictions are in fact "known interactions" present in either DrugBank or www.drugs.com. This fact has important implications on the significance of the validation of the predictions performed like the measurement of overlap of predictions and DDIs in FDA Adverse Event Reporting System, the associations of DDIs drugs with the same tissue and overlap with data of Tatonetti, et al. These evaluations should be done removing DDIs already present in DrugBank or www.drugs.com. Otherwise, it would not be clear whether the signal observed in the evaluations comes from the known DDIs or the novel predictions.
The method performs extremely well in cross-validation (as the authors acknowledge in the abstract), however the overlapping values obtained when comparing the predictions made using a dataset as training set with known interactions of a different dataset although significant, do not seem that impressive, which can be explained by data incompleteness in DDI knowledge but also by poor performance of the method in an external dataset. For this reason, I think it is important that the results of the validation of the predictions using other sources of potential DDIs are convincing.

The answer to the second point raises another question. What is PCRDs? In the manuscript is defined as "potential CYP-related DDIs" and it is not explained when it is used as validation. The authors should clarify when this set is used for validation in the text.

2nd Revision - authors' response 01 June 2012

Reviewer 1:

Perhaps the authors can describe in detail how they obtained this data so readers may also reproduce the study if they wished to.

_We now describe in detail the exact method by which we extracted the DDIs from Drugs.com as well as specify the database update version of the Drugs.com and DrugBank data (page 18). While we cannot provide the DDI training set due to restrictions posed by these sites, in order to allow replication of the cross-validation experiments, we also added the interactions and similarity measures with blinded drug IDs (Supplementary file S1)._}

Reviewer 2:

I am now confused about the meaning of "predictions" in this article. In the response to our second point, the authors acknowledge that a "significant portion of predictions intersect PCRDs- a set of known DDIs left for validation". Thus, this means that a high number of predictions are in fact "known interactions" present in either DrugBank or www.drugs.com. This fact has important implications on the significance of the validation of the predictions performed like the measurement of overlap of predictions and DDIs in FDA Adverse Event Reporting System, the associations of DDIs drugs with the same tissue and overlap with data of Tatonetti, et al. These evaluations should be done removing DDIs already present in DrugBank or www.drugs.com. Otherwise, it would not be clear whether the signal observed in the evaluations comes from the known DDIs or the novel predictions.

_The predictions defined as “novel” overlapping the FDA Adverse Event Reporting System did not include known DDIs. We now clarify it in the manuscript (page 13). We repeated other validations with the set of predictions that do not include known DDIs, obtaining similar significant results to those attained previously. Specifically, we repeated the evaluation against Tatonetti’s set, the tissue and co-morbidity validations (page 10) as well as the analysis of the electronic medical records data (page 13). We now clarify whenever the known DDIs are used for validation purposes._

The answer to the second point raises another question. What is PCRDs? In the manuscript is defined as "potential CYP-related DDIs" and it is not explained when it is used as validation. The authors should clarify when this set is used for validation in the text.

_PC RDs are known DDIs for which we could not easily identify their type - pharmacokinetic or pharmacodynamic. We clarify in the manuscript whenever this set is used for validation (page 9)._