



Concept Paper

In Silico Profiling of Clinical Phenotypes for Human Targets Using Adverse Event Data

Theodoros G. Soldatos , Guillaume Taglang and David B. Jackson *

Molecular Health GmbH, Kurfuersten Anlage 21, 69115 Heidelberg, Germany; soldatos@molecularhealth.com (T.G.S.); guillaume@taglang.org (G.T.)

* Correspondence: david.jackson@molecularhealth.com

Received: 23 October 2018; Accepted: 19 November 2018; Published: 23 November 2018



Abstract: We present a novel approach for the molecular transformation and analysis of patient clinical phenotypes. Building on the fact that drugs perturb the function of targets/genes, we integrated data from 8.2 million clinical reports detailing drug-induced side effects with the molecular world of drug-target information. Using this dataset, we extracted 1.8 million associations of clinical phenotypes to 770 human drug-targets. This collection is perhaps the largest phenotypic profiling reference of human targets to-date, and unique in that it enables rapid development of testable molecular hypotheses directly from human-specific information. We also present validation results demonstrating analytical utilities of the approach, including drug safety prediction, and the design of novel combination therapies. Challenging the long-standing notion that molecular perturbation studies cannot be performed in humans, our data allows researchers to capitalize on the vast tomes of clinical information available throughout the healthcare system.

Keywords: computational biology; large-scale approaches; systems pharmacology; adverse events; side-effects; outcome analytics; real world data; mode of action; clinical phenotypes; phenotypic screening; drug safety prediction

1. Introduction

Deciphering molecular mechanisms connecting disease phenotypes to the underlying genotypes remains one of the most important endeavors in science and medicine. Hidden within this molecular conundrum lies the knowledge that can power rational approaches to the detection, prevention and treatment of human disease. However, despite the enormous implications for our health and well-being, we still rely heavily on evidence from genetic and pharmacological perturbation experiments in model systems, such as cell-lines and model organisms. While certainly contributing to the identification of the key molecular protagonists of human disease, model systems suffer important limitations, the most fundamental of which remains their general lack of absolute congruence with the human condition that they are used to characterize [1]. While this is understandable given the enormous complexity and cell-type specificity of biological systems, it does suggest that new levels of therapeutic innovation might be achieved through development of methods that allow us to decipher molecular mechanisms directly from human specific data.

This is elegantly demonstrated by results from the TCGA project, where molecular data and phenotypic information have been analyzed to decipher novel targets and prognostic classifiers. Analysis of the TCGA endometrial carcinoma dataset, for example, has brought important new insights into the molecular nature of this disease [2], including the discovery of a new classification system based on four prognostically significant subgroups. Indeed, it can be argued that most recent clinical advancements emerge from analysis of patient-derived molecular data. However, vast tomes

of clinical information available throughout the healthcare system may remain underutilized, unless it becomes possible to derive molecular insights and hypotheses from these data.

One key challenge with using such sources of accessible human phenotype data, like electronic clinical reports, is the dearth of associated molecular data. We asked whether it might still be possible to connect a patient's clinical phenotype to molecular causation using the fundamental fact that physicians have been performing perturbation-studies on their patients for decades (Figure 1). From a molecular perspective, drugs are prescribed to modulate the function of disease-related molecular entities (typically proteins), with the goal of counteracting disease-causing mechanisms. Thus, it stands to reason that if we were able to link the molecular functions targeted by therapeutics to resultant side-effect phenotypes, and this for millions of patients, we should be able to computationally dissect the molecular mechanisms underlying these effects. This would enable production of datasets that might be of value to the molecular characterization of human disease directly from real world phenotype data.

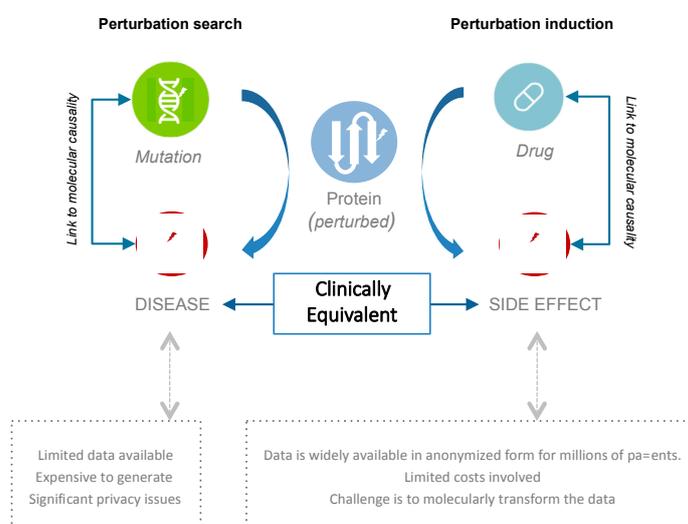


Figure 1. Perturbation studies are key to understanding human disease and side effects. Recent advances in high-throughput molecular technologies have revolutionized our ability to characterize the molecular foundations of human phenotypes. However, costs and privacy concerns remain a key impediment to the broad-scale analysis of such data. Our approach allows deciphering molecular mechanisms directly from accessible real-world clinical data: The phenotypic read-outs from drug interventions typically involve desired (e.g., cure) and/or undesired phenotypic outcomes (e.g., side-effects).

With this goal in mind, we developed a methodology for the molecular characterization of datasets providing drug-perturbation and phenotype observations. We analyzed data from the FDA Adverse Event Reporting System (FAERS) that contains millions of electronic adverse event (AE) case-reports, which document individual patients in de-identified format, their disease, and the side-effect consequences of co-reported drug treatments. While such data provides a direct connection between drugs and their phenotypic consequences for millions of patients, public resources such as the DrugBank [3] database are also available, which describe the targets of marketed and developmental drug therapies. It was therefore possible to connect phenotypes directly to targeted proteins/genes using a drug-centric data-integration process that maps drugs in each FAERS case-report, to their associated targets (Figure 2). Targets can further be contextualized by their pathway membership, permitting broader molecular connectivity with associated phenotypic outcomes. Using this approach, we extracted 1.8 million associations of clinical phenotypes to 770 human drug-targets.

In this work we describe our approach, the content of the produced phenotypic profiling of human targets, we present validation results, and give examples that demonstrate some of the key

analytical utilities. In specific, we discuss analytical modality concepts enabled by this approach, such as the exploration of clinical phenotypes caused by perturbation of human targets, the prediction of side-effects of new drugs, and the ability to perform virtual perturbation experiments by systematically comparing drug or target profiles. Overall, we find that insights derived directly from patient/clinical data can be more specific, compared to other methods.

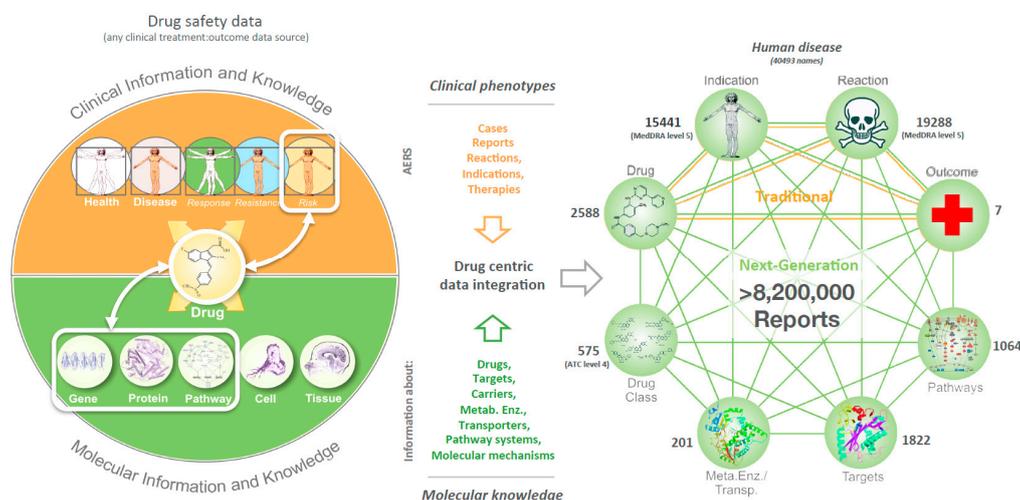


Figure 2. To enable the molecular analysis of drug induced clinical effects our approach adopts the simple premise that drugs induce side effects through perturbation of protein function. Drug-target knowledge can then be used to dissect the molecular basis of human disease. To achieve this, we built a drug-centric data integration process that currently combines observational clinical data for 8.2 million drug-safety reports with molecular information about drugs and their targets. The applied drug-centric data integration approach expands on traditional pharmacovigilance approaches to provide a new model for safety prediction and assessment through mechanism- and molecular-based analysis of drug induced phenotypes. Our target-reaction collection is a high dimensionality resource that to our knowledge contains more target-phenotype associations than any other relevant collection to-date. In specific, a multitude of 1.8 million target-reaction associations were characterized and can be explored via our dataset (Supplementary files 1, 2, 3, 4).

2. Materials and Methods

2.1. Connecting Clinical Phenotypes to Molecular Knowledge

Drugs typically modulate protein function to produce both desired (e.g., curative) and undesired (e.g., side-effects) clinical phenotypes. Given that side effects are essentially drug-induced diseases, we proposed that it would be possible to dissect the molecular basis of both side effects and diseases by analyzing phenotype data using drug-target knowledge. We achieved this via a drug-centric data integration process that combined clinical data for millions of patients with underlying molecular knowledge (targets, metabolizing enzymes, etc.). Results of this methodology are available as Supplementary files.

2.2. Data Integration

We developed a drug-centric data integration process that combines clinical data for about 6.8 million patients from FAERS (2000–2016), with the molecular knowledge in the DrugBank [3] database. This FAERS dataset holds about 8.2 million drug safety reports containing information about each patient's treatments, and disease indications and side-effects (termed reactions) coded using the MedDRA ontology [4]. The DrugBank database contains information about the molecular targets of marketed and developmental drugs. Matching treatments in each safety report to associated knowledge in DrugBank is complicated by text mining considerations [5], such as the non-standardized

drug names in FAERS. Key challenges lie in the fact that FAERS medication references may refer to one or to many drugs, may contain spelling errors or abbreviations, or may mention lump categories (e.g., vitamins, or antibiotics). We therefore developed a specialized mapping process that uses both drug synonym dictionaries and computational linguistic strategies to address these issues.

In terms of process, we employ a stepwise procedure. First, we link all FAERS names that are identical with a drug synonym. Then, for those medications that have not been mapped, we identify parts that are identical with or similar to a drug synonym, by using regular expressions. The process stops when the examined phrases differ by more than 5 characters. These features allow to compensate for common spelling variants or errors, but also to reduce the search space. In each phase of the mapping procedure a blacklist is used to exclude irrelevant terms (e.g., UNKNOWN, UNK) from being matched. By applying this mapping procedure, we successfully linked about 90% of all drug names mentioned in FAERS reports to single unambiguous therapeutic species, leading to at least one mapping for about 95% of all FAERS reports.

Further levels of molecular integration can be achieved by mapping all DrugBank proteins to their respective pathways within the Reactome [6,7] and PID (NCI-Nature & BioCarta) [8] resources. We have also integrated the ATC ontology from the WHO [9] to allow exploration and comparison of hierarchical drug classes. As a result, 8.2 million patient reports detailing drug-induced phenotypes could be linked to information about nine entity types: 15.4 K indications, 19.3 K reactions, seven clinical outcomes, and 2.6 K drugs, to 1.8 K targets, 201 metabolizing enzymes, 103 transporter proteins, 1 K pathways, and 881 ATC drug classes. Here we present results regarding targets, and reactions linked to at least 500 AE cases each (see Supplementary files 1, 2, 3, and 4).

2.3. Characterization of Relationships

Next, we sought to characterize the relative degree of statistical association between drugs and targets with reactions. For this purpose, we employed the proportional reporting ratio (PRR), an established measure of disproportionality in pharmacovigilance (Table 1). PRR gives an indication for the relative congruence of pairwise entity relations as based on their co-occurrence in subsets of AE data [10].

Table 1. For a drug (D) or a target (T) and an event (E) the PRR metric is defined as the value of $a(c + d)/c(a + b)$, based on the following contingency matrix.

AE Cases	Event (E)	Not E	Totals
D (T)	<i>a</i>	<i>b</i>	<i>A + b</i>
Not D (T)	<i>c</i>	<i>d</i>	<i>C + d</i>
Totals	<i>a + c</i>	<i>b + d</i>	<i>N = a + b + c + d</i>

Characterization of relationships in this way helps not only to identify known or novel relationships but also to perform relative comparisons. Furthermore, as the characterization of relationships spans all entities, a bidirectional study is possible: Physiological effects can be examined with respect to underlying mechanisms, or conversely, one can examine potential clinical results when specific molecular systems are perturbed. Moreover, previously reported events can be used to indicate (or predict) effects of potential therapies helping avoid erroneous prescriptions and toxic combinations.

2.4. The Collection of Relationships

We focused on 770 targets that were linked to at least 500 AE cases and were not listed in DrugBank [3] as also metabolizing enzymes, carriers, or transporters of other drugs. Phenotype co-occurrences of these ‘clean’ targets exceeded 12.9 million pairwise associations (i.e., target co-mentioning in AEs with each of seven clinical outcomes or indications and reactions expressed at any level of the MedDRA hierarchy). Inspecting phenotypes at higher levels of the MedDRA hierarchy helps compensate for occasional redundancy attributed to the reporting of reactions in AEs by using

similar terms or the reporting of reactions that are highly related (e.g., 'Coronary artery disease', 'Coronary artery thrombosis', and 'Coronary artery restenosis'). Association of targets to reaction classes of MedDRA level 2 alone gave rise to more than 230,000 relationships (Supplementary file 1). While PRR has become a standard metric for the evaluation of safety signals, we also considered relevant statistical significance be reflected by Fisher's exact test p -values and respective q -values [11]. For reference, we provide in Supplementary files 2, 3, and 4 all 1,601,362 co-occurrences of targets with reactions expressed at MedDRA level 4. Extracted targets and reactions were limited to those mentioned in at least 500 AEs each.

2.5. Presented Validation Examples

We also present validation results demonstrating analytical utilities of the approach, including drug safety prediction and the design of novel combination therapies.

2.5.1. Prospective Prediction of Side Effects

To validate the hypothesis, we used a dataset of older AEs that contained 2.4 million patients (AE cases from the 2000–2011 subset of FAERS). We examined adverse reactions and summarized them in terms of MedDRA level 1 categories. AE relationships were characterized using as measure the PRR metric.

2.5.2. Hypothesis Testing for Improved Use of Drugs

Beta-adrenergic receptor (BAR) activation affects processes involved in the progression of various cancer types [12,13]. We asked whether *BAR* antagonism could reduce skin cancer related mortality and therefore compared the clinical outcomes of two *BAR* activity modulations in a skin cancer patient population. The skin cancer cohort of this study comprised of those patients whose indication linked to the MedDRA (level 2) term 'Skin neoplasms malignant and unspecified'. Beta-blocking agents were defined with the help of the ATC classes C07AA, C07AB, and C07AG. Death occurrence was based on the reporting of the respective outcome. In absence of *BAR* antagonism, reported deaths occurred in 4059 out of 17,177 cases (23.6%), whereas with *BAR* inhibition deaths were reported in 241 out of 1308 cases (18.4%). The result for reduced skin cancer mortality with co-medication of beta-blockers is significant with Fisher's exact p -value = 0.000012.

2.6. Benchmarking

We relied on a benchmark set used previously in [14] and manually mapped 53 known protein–effect pairs [15] to our data. For the extraction we used the following criteria:

- (a) To ensure specificity we primarily considered reactions described at MedDRA level 4. In many cases AEs mentioned exactly a certain term as reported in the original dataset (e.g., 'Tachycardia'), whereas other times an effect could only be captured in AEs via higher MedDRA level descriptions (e.g., 'salivation' and 'Salivary gland disorders NEC'). In some other cases this strategy led to the identification of reaction terms too specific and with scant occurrence in AEs (e.g., 'Hypophosphatemia' or 'Hypocalcaemia').
- (b) We searched only for reactions: The physiological effects expected to occur when compounds hit the listed targets of [15] may be characterized with increased PRRs in AEs, however sometimes they may only reflect a condition and not an adverse reaction.
- (c) We searched for up to two synonyms per case: Sometimes the same effect may have been declared via its synonyms or in similar spelling (e.g., 'anemia' and 'anaemia'). In other cases, alternative names were searched to describe generic phrases of the original dataset (e.g., both 'hypertension' and 'hypotension' were searched for the originally termed effect 'blood pressure changes').
- (d) We also did not consider effects of proteins that were not targets or had no drugs reported in AEs.

The results of the manual mapping process can be found in Supplementary file 3, including cases that could not be exactly mapped (e.g., 'cell proliferation'), cases better described in higher levels (like 'Salivary gland disorders NEC' instead of 'salivation'), or cases when more than one terms may describe a condition. In those cases, phenotypes described in higher MedDRA levels would likely have been more appropriate to capture a certain valid association. This highlights the importance of accessing and assessing phenotype relationships at different levels/classes. Benchmarking challenges that could be attributed to the nature of FAERS reporting (i.e., the way terms are reported and coded) included redundancy and that a reaction might sometimes reflect an indication.

Overall, our search was manual and not exhaustive. Therefore, some of the selected terms may not be the most appropriate to highlight a certain target's relation to a phenotype as other synonyms may exist that have a better (higher/lower) signal. For the comparison task described next, we focused only on fifty derived heart and blood effects. We also kept three effects that were specific and with stronger signals, co-reported in more than 2500 AEs with respective targets (namely, 'Dry mouth', 'Electrocardiogram QT prolonged', and 'Dyskinesia').

2.7. Comparison of Target-Reaction Signals

We contrasted molecular analysis of phenotypes from AE patient prescriptions with methods that annotate proteins with side-effect information derived from phenotype associations of the drugs that target them. To represent the latter methods, we used the approach of [14] and for each reaction–target pair we counted the number of drugs that were co-reported with each effect, that target the protein, that do both, and neither. Then, we calculated respective PRRs and the p -value for overrepresentation using Fisher's exact test. The results indicate that some of the benchmarked relationships would have been dismissed by this approach (Supplementary Figure S1; Supplementary file 5). PRR-signals calculated directly from targets' AEs were generally more definitive which could be attributed to the fact that each drug individually may be mentioned in some AE observations but not (necessarily) in all cases that a certain target may have been perturbed. Our approach to characterize the molecular landscape of patient prescriptions allows thus extracting relationships more directly. To assess the variability of associations derived indirectly via drugs we listed in Supplementary file 5 also the respective maximum, average, and minimum drug-reaction PRR values, among the drugs targeting each target-reaction pair.

2.8. Data Availability

Phenotype annotations between targets and reactions are provided at Supplementary files 1, 2, 3, and 4.

3. Results

Our molecular characterization approach enables analysis and comparison of the phenotypic consequences associated with the perturbation of human proteins and/or to identify potential molecular mechanisms associated with distinct (or related) clinical phenotypes. It also enables the definition of patient-specific disease models in a manner that permits the dissection of underlying molecular mechanisms. In terms of numbers, FAERS (2000–2016) captured 15,441 indications (i.e., clinical phenotypes prior to drug treatment) and 19,288 AE reactions (i.e., clinical phenotypes after drug treatments).

3.1. Phenotypic Profiling of Human Target Proteins

Using drug-to-target data from DrugBank [3] and the treatment information provided for each patient in 8.2 million FAERS reports, 1822 human proteins were mapped to associated clinical phenotypes. Of these, 770 targets linked to at least 500 AE-cases and were listed in DrugBank [3] only as targets of drugs (and not as also metabolizing-enzymes, carriers, or transporters of other drugs). Quantifying the scope and diversity of phenotype observations (including indication and

reaction names expressed at any level of the MedDRA hierarchy, and outcomes), these 770 targets gave rise to more than 12.9 million target-phenotype associations (Supplementary Table S1). Described at MedDRA level-2 terms only, 333 reactions were mentioned in 230,841 combinations with these targets (Supplementary file 1). Of these associations, 155,820 (67.5%) were statistically significant ($q < 0.05$), while 42,284 (18.3%) connected via at least ten AE-cases and also have proportional reporting ratio (PRR) value ≥ 2 . Another 1,601,362 associations (Supplementary files 2, 3, and 4) between the 770 targets and 2803 reactions (MedDRA level-4 names) were characterized: 809,451 (50.5%) were found statistically significant ($q < 0.05$) and 290,440 (18.14%) appeared in at least ten AEs and have PRR ≥ 2 . These results highlight the specificity of many of the detected target-phenotype associations. They also emphasize the importance of exploring and assessing candidate associations at different levels of phenotype and target classes (Supplementary Figure S1).

Another aspect of this computational approach is that it enables systematic analysis of potential molecular players involved in human disease. The community can use our data to identify potential molecular protagonists across clinical phenotypes via a simple comparison strategy, permitting thus the step-by-step generation and dissection of molecular hypotheses.

3.2. Using the Approach: Examples and Perspectives of Analytical Strategies

Our approach allows linking drug-induced phenotypes to potential underlying molecular etiologies. Phenotypes can thus be analyzed and compared at different levels of phenotype and target classes (example at Supplementary Figure S1), at the level of any molecular perturbation (example at Supplementary Table S2), or specific clinical and molecular feature (examples at Supplementary Figures S2 and S3).

Importantly, comparative analytics permit the step-by-step generation and dissection of molecular hypotheses for assessing the functional association of targets with clinical phenotypes. For example, the most strongly associated reaction with *EGFR* perturbation was 'dermatitis acneiform' (extracted from Supplementary files 2, 3, and 4). This observation is consistent with current knowledge about the function of *EGFR* in epithelial biology [16–18] and demonstrates how even such a simple comparison strategy can help decipher phenotypic consequences of human target perturbation by using our resource.

Building on this principle we also demonstrated that it is possible to explore phenotypic consequences that result from different functional/activation states of a target (Supplementary Figure S1): We focused on a selection of targets whose function can be modulated in this way (activation versus antagonism) and determined reaction profiles with respect to the mechanism (agonism/inhibition) of the related drugs, using PRRs. Comparison of the phenotypic profiles reveals in each case differences between classes of targets and their functional modulation.

It is this, perhaps, one of the most important properties enabled by our approach for profiling individual patient prescriptions: The ability to define and investigate cohorts with specific characteristics (such as conditions, therapies, molecular entities) and compare them against other groups of patients. This allows performing systematically virtual perturbation experiments and identifying features or patterns by using a variety of methods, such as refined statistics, systems biology, and machine learning techniques. These perspectives do not directly compare to true perturbation studies (e.g., clinical trials, examination of side effects, experimental controls, etc.), but rather complement them. Real world application of our methodology has been shown in several instances to effectively help generate or validate hypotheses in terms of retrospective analysis of molecular or therapeutic properties e.g., [19,20].

3.3. Validation Results

To demonstrate some of the several predictive capabilities enabled by our approach we present two examples:

1. Prospective prediction of drug side effects. We asked whether it might be possible to predict AE-profiles of novel drugs. Examining the reaction profiles of tyrosine kinase inhibitors (TKIs) revealed that Sorafenib is more strongly associated with dermatological reactions than Sunitinib (Supplementary Table S2). Indeed, skin-related AEs are a common form of side-effects observed in patients treated with TKIs [16,17,21,22]. However, while these two TKIs share a common target-inhibition profile [23–26], Sorafenib additionally inhibits *BRAF*. We capitalized then on the Sorafenib-specific side-effect difference to hone in on the *BRAF*-specific clinical effects and hypothesized that the reaction profiles of other *BRAF* inhibitors may also include skin-related AEs. To test whether we could have predicted this relationship for the *BRAF*-specific inhibitor Vemurafenib, we used a data slice of all patient-cases reported prior to Vemurafenib's FDA-approval in August 2011. Through this prospective-retrospective analysis we found that *BRAF* perturbation was strongly related to dermatologic reactions (Supplementary Figure S2)—thus predicting the primary side-effects observed in Vemurafenib's phase-3 trial [27] (NCT01006980). Indeed, predicted skin-related AEs are consistent with *BRAF* pathway findings from recent studies and other independent work [28–31]. Importantly, this clinical effect is also included in Vemurafenib's FDA-label [32].
2. Rational prediction of combination therapies. We also assessed whether we could identify combinations of targets that may improve therapeutic outcomes in patients. We reasoned that AE outcomes and in particular “death rate”, might be used as surrogate marker for treatment efficacy. Using prior knowledge about the association between biobehavioral stress and tumorigenesis [12,13], we investigated whether perturbation of *beta-adrenergic receptor* (*BAR*) function in cancer patients might result in lower patient mortality. We examined the phenotypic consequences of *BAR* activity modulation by comparing two similar cohorts of skin cancer patients—one arm with and one without inhibition of *BAR* activity (Supplementary Figure S3). In absence of *BAR* antagonism, reported deaths occurred in 23.6% cases, whereas with *BAR* inhibition deaths were reported in 18.4% cases. These results suggested that co-medication of *BAR* blockers may be associated with reduced mortality of skin cancer patients, and were supported by subsequent in vitro and in vivo studies which demonstrated the role of *BAR* in tumor growth and stress-response signaling via *SRC* activation in cancer cells. Further analysis revealed that mortality was reduced across major cancer types in patients where *BAR* signaling is inhibited with beta-blockers and identified *BAR* inhibition as a potential combinatorial route to anti-cancer treatment [33]. In support of these observations, a recent phase-2 pilot-study (NCT01265576) examining clinical effects of beta-blocker usage in hepatocellular carcinoma (HCC) patients with VT-122 (combination of the non-selective beta-blocker Propranolol and the COX2-selective Etodolac) demonstrated potential anticancer effects when co-administered with Sorafenib by increasing therapy duration and overall survival, as compared to HCC patients treated with Sorafenib alone [34]. Our example study emphasizes the importance of clinical phenotype profiling at the level of target perturbations, but also provides a potentially novel approach to drug repositioning, via rational design of combination therapies.

3.4. Benchmarking and Comparison

We also benchmarked our findings against an independent dataset of known target–reaction effects, derived manually from [15] (Supplementary file 5). Overall, we find that characterization of target–reaction relationships are more conclusive when relying directly on patient/clinical data, as compared to methods that indirectly combine separate sources of drug side-effect data with molecular information e.g., [14,35]. Specifically, 56% more target–reaction relationships were recapitulated from the benchmark-set when we contrasted PRR-signals calculated over targets' AE-cases directly against those derived via respective drug–reaction associations. PRR-signals calculated directly from target AEs were generally more definitive, an effect likely attributable to the broader scope of these observations, as compared to those for each drug separately. In return, this allows identifying target-specific

phenotype effects that may otherwise remain masked, undermined by side-effect consequences of the combined co-perturbation of one drug's targets.

These results highlight the importance of direct characterization of target-phenotype relationships from large amounts of patient clinical data. However, an absolute comparison to similar methods that rely on other data is neither fair nor straightforward, and we believe that our results and results derived from different methods should be used in complementary ways. Also, our experience calls for the definition of a more proper benchmark to systematically compare similar approaches.

4. Discussion

Our dataset is the first and—to the best of our knowledge—currently only that permits molecular analysis of real-world phenotypes using accessible clinical data. It is also unique in that it enables scientists to quickly develop testable molecular hypotheses directly from human-specific observations, as opposed to model-organism data.

Previous systems-biology efforts to combine molecular information with phenotypes e.g., [14,36,37] relied primarily on side-effect information coming from labels e.g., [38,39]. AEs comprise an augmented source capturing real-world scenarios regarding drug use and combinations, phenotypes and conditions not studied in clinical-trials, and also include information for many more patients. However, FAERS data alone comes with limitations such as the lack of standardization [40]. Our work is an advance over previous target-phenotype association studies in that it addresses these shortcomings and provides a standardized approach to linking drug-induced AEs to potential underlying molecular etiologies. Specifically, with this work we combined previous efforts to, in addition, link phenotypes to the underlying molecular landscape, based on the annotation of AE patient-cases themselves. This characteristic of our approach permits phenotypes to be analyzed and compared at the level of any molecular perturbation or specific clinical and molecular feature (including drugs, drug-classes, targets, metabolizing enzymes, etc.). Furthermore, our strategy to characterize target-phenotype associations directly from their occurrence in clinical data, as opposed to using indirect drug-centric properties or observations e.g., [38,40], allows enriching existing integrative systems approaches with protein-perturbation datasets of additional levels of evidence [41,42].

Despite current limitations in known drug-target knowledge, we presented a number of case studies validating key utilities of our approach. We demonstrated how it could be used to identify potentially novel molecular players involved in diseases and illustrated how such molecular profiling can aid drug-safety prediction. Moreover, we demonstrated that it could be used to aid the identification of novel drug or target combinations associated with the improvement of specific phenotypes or outcomes. Importantly, comparison analytics over these results permit the step-by-step dissection of hypothesis for molecular causation and influence of individual targets on clinical phenotypes. In specific, our collection of target-phenotype associations can help identify targets associated with the exacerbation of specific phenotypes and/or an increase in their overall prevalence.

Integrating AE-data with molecular knowledge also expedites the identification of safety problems via a prospective approach. It provides drug-safety scientists with the unique ability to not only identify safety signals, but also assess whether any biomolecular rationale might exist to support traditional pharmacovigilance-based statistics. This is critical to both regulators and pharmaceutical companies, providing a more transparent assessment of whether a drug might be functionally involved in the emergence of a drug-safety issue. In this regard, our results continue to be utilized by clinical pharmacologists at the FDA in the process of drug relabeling due to newly identified safety issues [43,44].

In terms of future developments, we envisage a number of important data and analytical enhancements. We aim to combine AEs together with additional observational data and real-world evidence. We also expect that leveraging chemistry and genotype information will help towards the development of broader systems approaches. We hope that our work will provide additional context and scope to the efforts to interpret and model human disease and invite our dataset to serve as

foundation for an array of experiments, including analyses complementing data generated through direct characterization of patient genotypes.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2571-5135/7/4/37/s1>. Supplementary file 1: Association of targets to reaction classes of MedDRA level 2; Supplementary files 2, 3, and 4: Co-occurrences of targets with reactions expressed at MedDRA level 4; Supplementary file 5: Benchmark details and results; Supplementary file 6: Support tables and figures.

Author Contributions: D.B.J. conceived original idea; D.B.J. and T.G.S. prepared the manuscript; T.G.S. initiated and led the overall study; G.T. contributed with figures, analysis, and examples.

Funding: This research received no funding.

Acknowledgments: Molecular Health colleagues for advice and support.

Conflicts of Interest: D.B.J. and T.G.S. are currently employees of Molecular Health GmbH. G.T. was previously an employee of Molecular Health GmbH. T.G.S., G.T., and D.B.J. have patent applications associated with the analysis of adverse event data.

References

1. Horrobin, D.F. Modern biomedical research: An internally self-consistent universe with little contact with medical reality? *Nat Rev. Drug Discov.* **2003**, *2*, 151–154. [[CrossRef](#)] [[PubMed](#)]
2. Cancer Genome Atlas Research Network; Kandoth, C.; Schultz, N.; Cherniack, A.D.; Akbani, R.; Liu, Y.; Shen, H.; Robertson, A.G.; Pashtan, I.; Shen, R.; et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **2013**, *497*, 67–73.
3. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672. [[CrossRef](#)] [[PubMed](#)]
4. MedDRA. Available online: <http://meddra.org/> (accessed on 21 March 2017).
5. Soldatos, T.G.; Perdigão, N.; Brown, N.P.; Sabir, K.S.; O'Donoghue, S.I. How to learn about gene function: Text-mining or ontologies? *Methods* **2015**, *74*, 3–15. [[CrossRef](#)] [[PubMed](#)]
6. Croft, D.; Mundo, A.F.; Haw, R.; Milacic, M.; Weiser, J.; Wu, G.; Caudy, M.; Garapati, P.; Gillespie, M.; Kamdar, M.R.; et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **2014**, *42*, D472–D477. [[CrossRef](#)] [[PubMed](#)]
7. Milacic, M.; Haw, R.; Rothfels, K.; Wu, G.; Croft, D.; Hermjakob, H.; D'Eustachio, P.; Stein, L. Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers* **2012**, *4*, 1180–1211. [[CrossRef](#)] [[PubMed](#)]
8. Schaefer, C.F.; Anthony, K.; Krupa, S.; Buchoff, J.; Day, M.; Hannay, T.; Buetow, K.H. PID: The Pathway Interaction Database. *Nucleic Acids Res.* **2009**, *37*, D674–D679. [[CrossRef](#)] [[PubMed](#)]
9. WHOCC—Structure and Principles. Available online: https://www.whocc.no/atc/structure_and_principles/ (accessed on 21 March 2017).
10. Van Puijenbroek, E.P.; Bate, A.; Leufkens, H.G.; Lindquist, M.; Orre, R.; Egberts, A.C. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol. Drug Saf.* **2002**, *11*, 3–10. [[CrossRef](#)] [[PubMed](#)]
11. Storey, J.D.; Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9440–9445. [[CrossRef](#)] [[PubMed](#)]
12. Thaker, P.H.; Han, L.Y.; Kamat, A.A.; Arevalo, J.M.; Takahashi, R.; Lu, C.; Jennings, N.B.; Armaiz-Pena, G.; Bankson, J.A.; Ravoori, M.; et al. Chronic stress promotes tumor growth and angiogenesis in a mouse model of ovarian carcinoma. *Nat. Med.* **2006**, *12*, 939–944. [[CrossRef](#)] [[PubMed](#)]
13. Antoni, M.H.; Lutgendorf, S.K.; Cole, S.W.; Dhabhar, F.S.; Sephton, S.E.; McDonald, P.G.; Stefanek, M.; Sood, A.K. The influence of bio-behavioural factors on tumour biology: Pathways and mechanisms. *Nat. Rev. Cancer* **2006**, *6*, 240–248. [[CrossRef](#)] [[PubMed](#)]
14. Kuhn, M.; Al Banchaouchi, M.; Campillos, M.; Jensen, L.J.; Gross, C.; Gavin, A.C.; Bork, P. Systematic identification of proteins that elicit drug side effects. *Mol. Syst. Biol.* **2013**, *9*, 663. [[CrossRef](#)] [[PubMed](#)]
15. Whitebread, S.; Hamon, J.; Bojanic, D.; Urban, L. Keynote review: In vitro safety pharmacology profiling: An essential tool for successful drug development. *Drug Discov. Today* **2005**, *10*, 1421–1433. [[CrossRef](#)]
16. Perez-Soler, R. Rash as a surrogate marker for efficacy of epidermal growth factor receptor inhibitors in lung cancer. *Clin. Lung Cancer* **2006**, *8* (Suppl. 1), S7–S14. [[CrossRef](#)] [[PubMed](#)]

17. Liu, H.; Wu, Y.; Lv, T.F.; Yao, Y.W.; Xiao, Y.Y.; Yuan, D.M.; Song, Y. Skin Rash could Predict the Response to EGFR Tyrosine Kinase Inhibitor and the Prognosis for Patients with Non-Small Cell Lung Cancer: A Systematic Review and Meta-Analysis. *PLoS ONE* **2013**, *8*, e55128. [CrossRef] [PubMed]
18. Kozuki, T. Skin problems and EGFR-tyrosine kinase inhibitor. *Jpn. J. Clin. Oncol.* **2016**, *46*, 291–298. [CrossRef] [PubMed]
19. Racz, R.; Soldatos, T.G.; Jackson, D.; Burkhart, K. Association between Serotonin Syndrome and Second-Generation Antipsychotics via Pharmacological Target-Adverse Event Analysis. *Clin. Transl. Sci.* **2018**. [CrossRef] [PubMed]
20. Soldatos, T.G.; Dimitrakopoulou-Strauss, A.; Larribere, L.; Hassel, J.C.; Sachpekidis, C. Retrospective Side Effect Profiling of the Metastatic Melanoma Combination Therapy Ipilimumab-Nivolumab Using Adverse Event Data. *Diagnostics* **2018**, *8*, 76. [CrossRef] [PubMed]
21. Poprach, A.; Pavlik, T.; Melichar, B.; Puzanov, I.; Dusek, L.; Bortlicek, Z.; Vyzula, R.; Abrahamova, J.; Buchler, T. Skin toxicity and efficacy of sunitinib and sorafenib in metastatic renal cell carcinoma: A national registry-based study. *Ann. Oncol.* **2012**, *23*, 3137–3143. [CrossRef] [PubMed]
22. Zheng, Y.; Wang, F.; Wu, G.; Zhang, L.; Wang, Y.; Wang, Z.; Chen, P.; Wang, Q.; Lu, J.; Wang, Y.; et al. The Relationship between the Adverse Events and Efficacy of Sorafenib in Patients with Metastatic Renal Cell Carcinoma. *Medicine* **2015**, *94*, e2222. [CrossRef] [PubMed]
23. Van Erp, N.P.; Gelderblom, H.; Guchelaar, H.J. Clinical pharmacokinetics of tyrosine kinase inhibitors. *Cancer Treat. Rev.* **2009**, *35*, 692–706. [CrossRef] [PubMed]
24. Di Gion, P.; Kanefendt, F.; Lindauer, A.; Scheffler, M.; Doroshenko, O.; Fuhr, U.; Wolf, J.; Jaehde, U. Clinical pharmacokinetics of tyrosine kinase inhibitors: Focus on pyrimidines, pyridines and pyrroles. *Clin. Pharmacokinet.* **2011**, *50*, 551–603. [CrossRef] [PubMed]
25. Sorafenib: FDA Label—NEXAVAR (in DailyMed). Available online: <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=b50667e4-5ebc-4968-a646-d605058dbef0> (accessed on 29 April 2016).
26. Sunitinib: FDA Label—SUTENT (in DailyMed). Available online: <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=43a4d7f8-48ae-4a63-9108-2fa8e3ea9d9c> (accessed on 29 April 2016).
27. Chapman, P.B.; Hauschild, A.; Robert, C.; Haanen, J.B.; Ascierto, P.; Larkin, J.; Dummer, R.; Garbe, C.; Testori, A.; Maio, M.; et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **2011**, *364*, 2507–2516. [CrossRef] [PubMed]
28. Blank, C.U.; Larkin, J.; Arance, A.M.; Hauschild, A.; Queirolo, P.; Del Vecchio, M.; Ascierto, P.A.; Krajsova, I.; Schachter, J.; Neyns, B.; et al. Open-label, multicentre safety study of vemurafenib in 3219 patients with BRAFV600 mutation-positive metastatic melanoma: 2-year follow-up data and long-term responders' analysis. *Eur. J. Cancer* **2017**, *79*, 176–184. [CrossRef] [PubMed]
29. Kopetz, S.; Desai, J.; Chan, E.; Hecht, J.R.; O'Dwyer, P.J.; Maru, D.; Morris, V.; Janku, F.; Dasari, A.; Chung, W.; et al. Phase II Pilot Study of Vemurafenib in Patients with Metastatic BRAF-Mutated Colorectal Cancer. *J. Clin. Oncol.* **2015**, *33*, 4032–4038. [CrossRef] [PubMed]
30. Huang, V.; Hepper, D.; Anadkat, M.; Cornelius, L. Cutaneous toxic effects associated with vemurafenib and inhibition of the BRAF pathway. *Arch. Dermatol.* **2012**, *148*, 628–633. [CrossRef] [PubMed]
31. Su, F.; Viros, A.; Milagre, C.; Trunzer, K.; Bollag, G.; Spleiss, O.; Reis-Filho, J.S.; Kong, X.; Koya, R.C.; Flaherty, K.T.; et al. RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. *N. Engl. J. Med.* **2012**, *366*, 207–215. [CrossRef] [PubMed]
32. Vemurafenib: FDA Label—ZELBORAF (in DailyMed). Available online: <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=38eea320-7e0c-485a-bc30-98c3c45e2763> (accessed on 14 April 2016).
33. Armaiz-Pena, G.N.; Allen, J.K.; Cruz, A.; Stone, R.L.; Nick, A.M.; Lin, Y.G.; Han, L.Y.; Mangala, L.S.; Villares, G.J.; Vivas-Mejia, P.; et al. Src activation by β -adrenoreceptors is a key switch for tumour metastasis. *Nat. Commun.* **2013**, *4*, 1403. [CrossRef] [PubMed]
34. De la Torre, A.N.; Castaneda, I.; Hezel, A.F.; Bascomb, N.F.; Bhattacharyya, G.S.; Abou-Alfa, G.K. Effect of coadministration of propranolol and etodolac (VT-122) plus sorafenib for patients with advanced hepatocellular carcinoma (HCC). *J. Clin. Oncol.* **2015**, *33*. [CrossRef]
35. Luo, Y.; Zhao, X.; Zhou, J.; Yang, J.; Zhang, Y.; Kuang, W.; Peng, J.; Chen, L.; Zeng, J. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* **2017**, *8*, 573. [CrossRef] [PubMed]

36. Kuhn, M.; Campillos, M.; González, P.; Jensen, L.J.; Bork, P. Large-scale prediction of drug-target relationships. *FEBS Lett.* **2008**, *582*, 1283–1290. [[CrossRef](#)] [[PubMed](#)]
37. Campillos, M.; Kuhn, M.; Gavin, A.C.; Jensen, L.J.; Bork, P. Drug target identification using side-effect similarity. *Science* **2008**, *321*, 263–266. [[CrossRef](#)] [[PubMed](#)]
38. Kuhn, M.; Letunic, I.; Jensen, L.J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **2016**, *44*, D1075–D1079. [[CrossRef](#)] [[PubMed](#)]
39. Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L.J.; Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **2010**, *6*, 343. [[CrossRef](#)] [[PubMed](#)]
40. Banda, J.M.; Evans, L.; Vanguri, R.S.; Tatonetti, N.P.; Ryan, P.B.; Shah, N.H. A curated and standardized adverse drug event resource to accelerate drug safety research. *Sci. Data* **2016**, *3*, 160026. [[CrossRef](#)] [[PubMed](#)]
41. Boland, M.R.; Jacunski, A.; Lorberbaum, T.; Romano, J.D.; Moskovitch, R.; Tatonetti, N.P. Systems biology approaches for identifying adverse drug reactions and elucidating their underlying biological mechanisms. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2016**, *8*, 104–122. [[CrossRef](#)] [[PubMed](#)]
42. Wang, Z.; Clark, N.R.; Ma'ayan, A. Drug-induced adverse events prediction with the LINCS L1000 data. *Bioinformatics* **2016**, *32*, 2338–2345. [[CrossRef](#)] [[PubMed](#)]
43. Duggirala, H.J.; Topping, J.M.; Smith, E.; Bright, R.A.; Baker, J.D.; Ball, R.; Bell, C.; Bright-Ponte, S.J.; Botsis, T.; Bouri, K.; et al. Use of data mining at the Food and Drug Administration. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 428–434. [[CrossRef](#)] [[PubMed](#)]
44. Health Informatics at FDA > Data Mining at FDA—White Paper ‘Summary of Past and Present Data Mining Activities at the Food and Drug Administration’. Available online: <http://www.fda.gov/scienceresearch/healthinformatics/ucm446239.htm> (accessed on 21 January 2016).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).