

Exploring the Potential of Direct-To-Consumer Genomic Test Data for Predicting Adverse Drug Events

Patrick M. Zhang, Indra Neil Sarkar, PhD, MLIS
Center for Biomedical Informatics, Brown University, Providence, RI

Abstract

Recent technological advancements in genetic testing and the growing accessibility of public genomic data provide researchers with a unique avenue to approach personalized medicine. This feasibility study examined the potential of direct-to-consumer (DTC) genomic tests (focusing on 23andMe) in research and clinical applications. In particular, we combined population genetics information from the Personal Genome Project with adverse event reports from AEOLUS and pharmacogenetic information from PharmGKB. Primarily, associations between drugs based on co-occurring genetic variations and associations between variants and adverse events were used to assess the potential for leveraging single nucleotide polymorphism information from 23andMe. The results of this study suggest potential clinical uses of DTC tests in light of potential drug interactions. Furthermore, the results suggest great potential for analyzing associations at a population level to facilitate knowledge discovery in the realm of predicting adverse drug events.

Introduction

Adverse drug events (ADEs) are a significant challenge facing global public health. In addition to causing clinical harm to patients that can range from a minor headache to death, ADEs represent an economic burden in the form of hospital admissions, prolonged stays, and additional treatments. Although exact figures vary between sources, the estimated annual cost in the United States alone is \$30.1 billion, where each incident can cost around \$3,000 depending on severity^{1,2}. Spontaneous reporting systems are valuable resources for postmarket pharmacovigilance, presenting collected reports of ADEs and medication errors that are voluntarily submitted by clinicians, healthcare facilities, and patients. The US Food and Drug Administration (FDA) adverse event reporting system (FAERS) is one such database. As with all spontaneous reporting systems, limitations exist for FAERS; a report for a drug and adverse effect does not necessarily demonstrate a causal relationship between them, and not all adverse events for a particular drug may be reported³. A recently curated and standardized version of FAERS has been made publicly available, called the Adverse Event Open Learning through Universal Standardization (AEOLUS). AEOLUS provides standardized data and correlative statistics about drugs administered for an indication and the adverse outcomes based on data originating in FAERS⁴.

While an estimated 50% of ADEs are preventable and result from a medication error, many are due to genetic variations that lead to a heightened reaction to a drug. Pharmacogenetics focuses on the study of pharmacology in the context of genomics, aiming to develop therapies that maximize efficacy and minimize adverse drug events. The Pharmacogenomics Knowledge Base (PharmGKB)⁵ is a resource developed by the Pharmacogenomics Research Network that facilitates exploring the effect of genetic variation on drug response. The data contained within PharmGKB is the product of utilizing natural language processing techniques on pharmacogenomic articles from PubMed and verification through manual curation. The vocabulary for reference single nucleotide polymorphism cluster IDs (RSIDs) and drugs are primarily standardized through dbSNP⁶ and DrugBank⁷. An annotation for a given variant is included if a peer-reviewed article contains an association between a gene, drug, and disease. This study focuses on variants associated with an increased risk of an adverse event in response to a drug.

The role of pharmacogenomic tests in clinical practice has been expanding, with recent advances in technology that make genome-wide studies both economically and practically feasible. Several publications have demonstrated the potential clinical utility of pharmacogenetic tools, which reduced re-hospitalizations and lowered health care costs by 84% compared to controls^{8,9}. Clinical applications of personalized medicine are partially limited by the current shortcomings of genome-wide association studies that include difficulty obtaining large sample sizes, particularly for minority populations¹⁰. Short of genomic testing being available and affordable to inform every clinical

encounter, direct-to-consumer (DTC) genetic testing is an emerging technology that has the potential to fill the gap of available genomic data. For instance, 23andMe is a leading producer of DTC tests and provides consumers with their genetic information without the need for a healthcare professional. This information includes inherited variants associated with risk factors for conditions and hypersensitivity to drugs. A recent study concluded that 23andMe's tests generally mirror clinically relevant pharmacogenetics tests; however, limitations included inconsistencies in having non-specific tests and lacking relevant ones, as well as the difficulty of addressing differences in variant frequencies between ethnicities¹¹. For that reason, 23andMe notes that its product is not intended to be a diagnostic tool, but the wealth of information that this accessible technology can provide may have promising utility for research purposes. 23andMe variant profiles for 768 individuals are publicly available at the Harvard Personal Genome Project (PGP)¹². PGP is a collection of public genomes from volunteers that consent to making their non-anonymous variant data public for research.

This feasibility study focused on two objectives. The first was to examine the clinical relevance of DTC reports through a thorough exploration of the population data. While not diagnostic, 23andMe data might serve as an additional source of patient-supplied information for supporting clinical decisions and prompt additional, clinical-grade genetic tests when necessary. The secondary focus of this study was to discover pharmacogenetic associations to guide future research directions or to act as a separate resource by creating a networks of drugs and adverse events. The approach is outlined in Figure 1. Briefly, 23andMe variant profiles were examined from a population perspective to determine the prevalence of PharmGKB-associated annotated variants and consequently to find associated drugs. Then, RSIDs were directly mapped to ADEs by combining the resources provided by PharmGKB and AEOLUS, which may be a source for possible variant-outcome associations despite lacking explicit genetic information.

Methods

PharmGKB: Variant-drug Associations

Only variant-drug associations previously identified to increase the risk for ADEs were included in this study. The Variant and Clinical Associations file from PharmGKB includes this information and refers to the type of association as phenotype category. The categories were either one or combinations of the following: toxicity, efficacy, dosage, metabolism/PK, and other. Each variant-drug pair additionally contained a summary of the publication's findings. Annotations involving RSIDs were filtered for significance, and the annotations were divided into those that indicated increases and decreases in the incidence of a particular event. First, only the toxicity associations were examined. Language patterns of the summaries were analyzed and a series of regular expressions were devised to obtain a vocabulary for relevant outcomes; for example, outcomes usually were found between the phrases "risk/severity/likelihood of" and a conditional word like "when." Outcomes like "dose reduction" or "non-response" were examples of those mislabeled as toxicity. This vocabulary was then applied to classify toxicity annotations within other categories.

Next, the effect of genetic variation on a drug was summarized. In particular, annotations were examined to determine whether variants on different chromosomes could influence sensitivity for a drug. For this study, a group of 42 drugs that treat mental health or psychiatric disorders was selected and obtained from Drugs.com¹³. The Variant and Clinical Associations file was joined with PharmGKB's RSID file, which includes variant information from dbSNP such as location. Weight gain and neutropenia were selected as outcomes affected by variants on multiple chromosomes that occurred frequently for mental health drugs and for a wider range of drugs, respectively. The genes of annotations for these adverse events were compared to explore genetic variation and drug sensitivity.

Exploration of PGP Patient Profiles

PGP reports were downloadable as text files, containing data in columns corresponding to the RSID, chromosome, position number, and affected alleles. Less common file types like vcf and bam were converted to csv or txt when appropriate. In cases where participants uploaded multiple files, the most recent text file was used. In some cases, the more recent profiles had included fewer variants than a previous one, so in these situations the original upload was used. Six participants uploaded reports from Gene for Good, and five reports were generated from AncestryDNA, which were selected only if standard 23andMe profiles were not available. To better describe the population of reports from PGP, the contents were compared. The similarity between reports was determined by

evaluating the total number of contained variants and their identities. Matching reports refer to two reports where the variants in the smaller report is entirely contained in the larger one. Reports that match may belong to a subset within the population. Lastly, the reports were used to explore the co-occurrence of RSIDs on different chromosomes. The presence of variants associated with an increased risk of weight gain and neutropenia were checked within a selected portion of the total reports.

Using PGP to Find Associated Variants and Drugs Within a Population

Variants commonly present in individuals within a population were found in this study by matching the 23andMe patient profiles to the annotated PharmGKB RSIDs. The result was displayed in a binary table with columns representing the RSIDs and rows representing participant files. From this, the occurrences of RSIDs with a particular variant were counted; this was achieved by finding the subset of reports containing a particular variant and counting the number of these reports that contained each of the other RSIDs. Thus each variant was represented by a subset of all 23andMe profiles, and RSIDs were considered related in that subset if they were shared by 90% of the subset, an arbitrary significance level. Overall association between variants was determined by examining all subgroups.

Briefly, for each annotated RSID, the related variants were found. Within these related variants, those that had annotations for the drug of interest were obtained, and lastly with these variants, the occurrence of all drugs were counted. The fraction of RSIDs associated with a given drug was calculated; a value of 1 indicated that the drug was associated with all RSIDs and should be closely related to the drug of interest. Since the type of association between drugs obtained using this methodology is unknown, multiple means of validation were necessary. DrugBank includes a list of interacting drugs, which was compared to the drugs indicated by genetic association. AEOLUS provides similar information on interactions in the form of case reports. Within a case, a drug may be considered a primary suspect, secondary suspect, concomitant, or interacting drug. For cases where the drug of interest was the primary suspect, all interacting drugs were obtained. Interacting drugs were also considered to be primary drugs in cases where the drug of interest was labeled as an interacting drug. A similar process may be applied to secondary suspect and concomitant drugs, or those that are typically administered in conjunction.

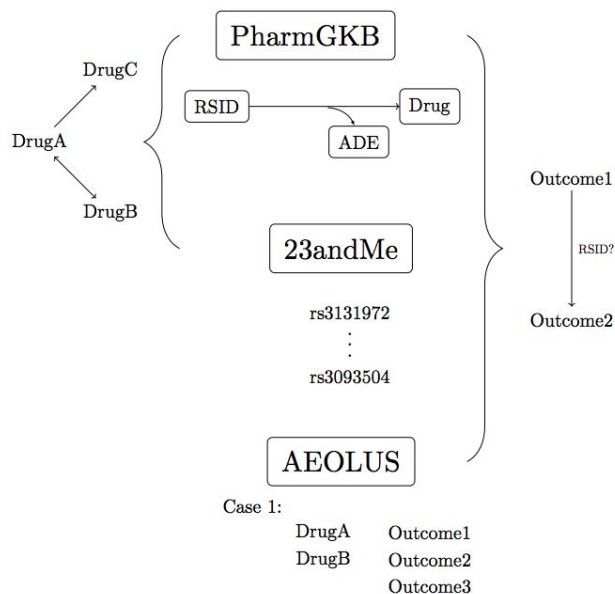


Figure 1. Schematic summary of the proposed methodology. Pharmacogenomic information (PharmGKB) was combined with population genetic data (23andMe) to find associations between drugs. Reports of ADEs or outcomes (AEOLUS) were combined with 23andMe data to find associated outcomes with potential genetic associations.

Mapping Variants to Adverse Events

Adverse event information from PharmGKB and AEOLUS was used to map variants to adverse events. AEOLUS contains reports of ADEs without indications of genetic association. For a particular drug, outcomes from all cases in which it was the primary suspect and only administered drug were examined. Edges were drawn between outcomes within a case and represented as pairs of associated ADEs, and the total occurrence of these pairs across all cases reflected the degree of correlation. These relationships may have contained outcomes included in PharmGKB associations, which facilitated a partial mapping of RSIDs directly to ADEs. These annotated ADEs were manually curated for the drug of interest from the annotations file and standardized to the vocabularies used in the AEOLUS database.

Gephi Visualization

Gephi is a visualization software for networks¹⁴. The application has several force-directed layout algorithms to distribute nodes and edges, and the layouts used in each graph were chosen to most clearly display clusters and associations. The Fruchterman-Reingold layout models nodes and edges with attractive and repulsive forces and distributes them such that the overall energy of the system is minimized. In the force atlas layout, nodes and edges similarly repulse and attract, but the layout favors bringing poorly connected nodes closer to very connected nodes. Lastly, the Yifan Hu layout combines standard methods but treats clusters as a single node when determining repulsive forces. Gephi offers several statistical measures to assist in quantifying node associations. The modularity of a network measures the connectedness of a graph and its calculation involves comparing the density of edges within a community versus links between communities. Communities are indicative of some form of association¹⁵.

Relationships between drugs were visualized using Gephi with edge weights represented by the ratio of associated RSIDs. The classes of drugs were defined using PubChem¹⁶ and the standard anatomical therapeutic chemical (ATC) classification system obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG)¹⁷. The nodes were colored by drug class or by modularity to best display associations; the node sizes were proportional to its degree, or number of incoming and outgoing edges. Associations between ADEs were visualized using a similar approach, where nodes and edge weights represented outcomes and the correlations between them. The ADEs found in PharmGKB annotations were labeled on the visualization, and the results were filtered by edge weight. The filtered network utilized the Yifan Hu layout in combination with the force-directed algorithms. Similarly, the nodes were colored by modularity and sized proportional to its weighted degree.

Results

Variant-Drug Annotations

Using the PharmGKB annotations file, 2675 variant-drug pairs were obtained along with the outcome, capturing 994 unique RSIDs and 170 unique drugs. Table 1 below summarizes shared variant-drug associations for 26 of the 42 mental health drugs, grouped by classifications curated from PubChem.

DTC Genomic Reports

699 DTC reports were downloaded from 23andMe or other sources and used for the analysis. Complete genome tests were desired, so uploaded Y-chromosome or mitochondrial DNA tests were removed. The number of RSIDs in each report ranged from 546,058 to 1,003,774. 836 out of the 994 (84.1%) annotated RSIDs were found within the reports. Out of the 283 unique annotated drugs, 274 (93.5%) were contained. The median number of RSIDs in the total reports was 710. When disregarding the genotype for each variant, thirty-four unique sets of variants were present among the total reports. The largest two sets were composed of 710 and 581 variants and encompassed 281 and 127 participant reports, respectively. To explore the co-occurrence of variants on different chromosomes, the presence of variants associated with increased weight gain was checked within the two sets. The contents were identical except for one missing variant in the smaller set. Each of the ten drugs with an increased risk for weight gain were affected by variants on multiple chromosomes, ranging between two and seventeen total RSIDs. Two RSIDs on different chromosomes, which affected three of the drugs, were not found in the reports.

Table 1. Overlapping variants for atypical antipsychotics (except haloperidol, which is a typical antipsychotic), anticonvulsants (AC), selective serotonin reuptake inhibitors (SSRI), and tricyclic antidepressants (TCA). The presence of a variant-drug annotation is denoted by a dot.

		rs489693	rs3813929	rs324420	rs17782313	rs518147	rs1414334	rs3780412	rs3780413
Antipsychotics	haloperidol*	•	•	•		•			
	aripiprazole	•		•					
	clozapine	•	•	•	•	•	•	•	•
	iloperidone		•						
	olanzapine	•	•	•	•	•	•	•	•
	paliperidone	•			•				
	quetiapine	•	•	•	•				
	risperidone	•	•	•	•	•	•	•	•
	ziprasidone	•	•						
AC	carbamazepine	•	•	•	•	•	•	•	•
	phenytoin	•	•	•	•	•	•	•	•
	valproic acid	•							
SSRI	citalopram						•	•	
	escitalopram								
	fluvoxamine						•	•	
	nefazodone			•	•	•			
	ondansetron			•	•				
	paroxetine			•	•	•	•	•	•
	sertraline						•	•	
	venlafaxine			•	•	•	•	•	
TCA	amitriptyline	•							
	clomipramine	•	•	•	•				
	doxepin	•							
	imipramine	•							
	nortriptyline	•							

Associated Drugs

Drugs that treat mental illnesses were chosen for this analysis. The network of drug associations is shown in Figure 2a, where the nodes represent drugs and are colored by modularity clusters. The labeled nodes are those identified from Drugs.com, and the non-labeled are those that were only present in PharmGKB annotations. Therefore, non-labeled drugs connected by edges or in the same cluster are those with a possible association determined through shared RSIDs. For example, the labeled nodes within the green community are exclusively selective serotonin reuptake inhibitor (SSRI) drugs (escitalopram, citalopram, fluvoxamine, and sertraline) and phenothiazine antipsychotics (chlorpromazine, fluphenazine, thioridazine, and trifluoperazine). Non-labeled nodes include the drug class antipsychotics, which supports the validity of the clustering. Milnacipran, another unlabeled drug, is an antidepressant in the serotonin-norepinephrine reuptake inhibitor class, and although it was not included the original list of drugs, its enantiomer levomilnacipran was. Other nodes in the cluster were caffeine, fenofibrate, terbinafine, and ticlopidine. The latter three drugs were only connected to sertraline in the cluster.

Paroxetine was chosen to perform a closer analysis on associated drugs for a specific use case. The resulting network around paroxetine is shown in Figure 2b, and these associations are those that are based on only shared variants from 23andMe. 34 drugs were found to be related, in addition to two drug classes, antipsychotics and taxanes. One reason for association may be drug interactions, and DrugBank provides an extensive list of interactions obtained from drug labels and scientific literature. For paroxetine, DrugBank lists 779 interacting drugs, although a majority of them cite outcomes that were excluded from this study such as a decrease in serum concentration of a drug. 30 out of the 34 23andMe predicted drug interactions were listed as interacting drugs in DrugBank, and five of these had outcomes that matched the types of toxicities reported in the PharmGKB annotations. The drugs associated to paroxetine were also compared to those found in reported cases from AEOLUS. 24 drugs were obtained, and every drug was listed among the interacting drugs on DrugBank, while 13 of them had toxicity-related outcomes. There was no overlap between associated drugs found through AEOLUS and 23andMe genetic reports.

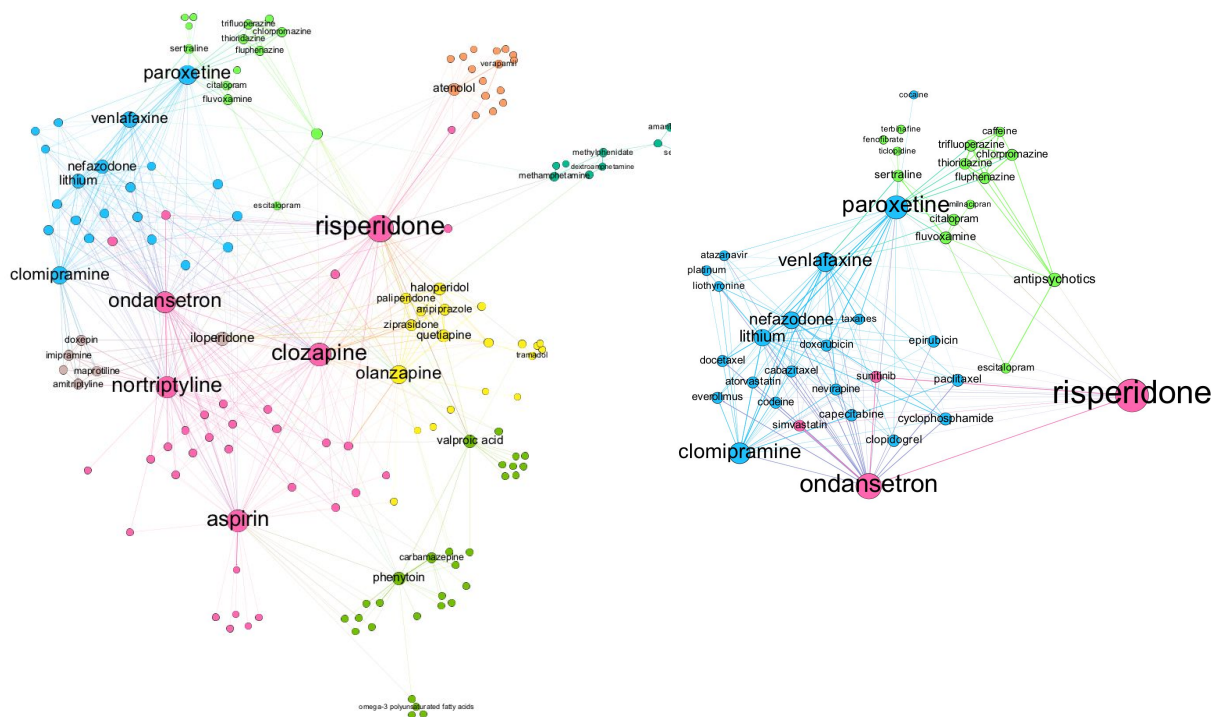


Figure 2. Network of drug associations for those that treat psychiatric disorders: **(a)** (left) An overall network that was distributed using a force-directed algorithm, and the nodes were colored by modularity to indicate clusters; **(b)** (right) Closer figure of the network of drugs associated with paroxetine. The green cluster is primarily composed of antipsychotics and SSRIs.

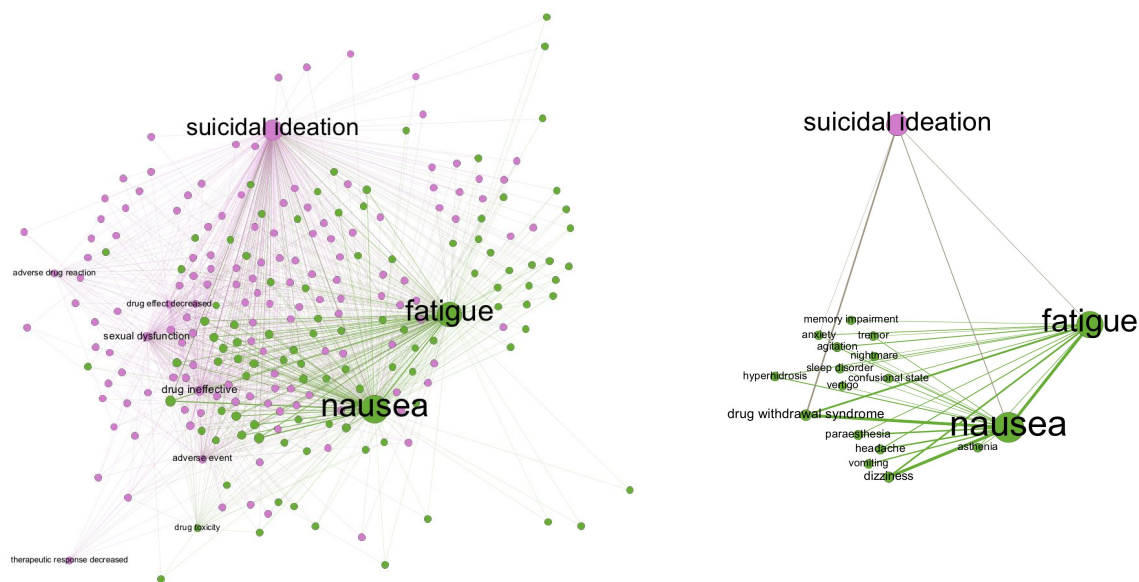


Figure 3. Adverse events associated with Paroxetine: **(a)** (left) Network of ADEs reported in cases where paroxetine was the only drug administered. The distribution was the result of a force-directed algorithm, and the nodes were colored by modularity and sized proportional to the weighted degree. **(b)** (right) Displays the network of ADEs related to nausea, fatigue, and suicidal ideation.

Variant-ADE Associations

Adverse events for paroxetine were examined within the AEOLUS database, and the resulting associations are graphed in Figure 3a. ADEs with a PharmGKB variant annotation were labeled; nausea, fatigue, and suicidal ideation were found to occur the most frequently and had the greatest weighted degrees. The network was filtered by edge weight to limit the associations to those connected to one or more of the three outcomes (Figure 3b). While many of these ADEs were linked to both nausea and fatigue, two communities are distinguishable. Notably, dizziness, vomiting, and headache were clustered with nausea; and tremors, vertigo, and confusional state were grouped with fatigue. From these results, there was no clear relationship between the variants. Nausea and vomiting each had one variant RSID; patients with rs762551 had an increased risk of fatigue, and the variant is located on chromosome 15, affecting CYP1A2. rs1176744 was associated with discontinuation syndrome and nausea, and it is located on chromosome 11, affecting HTR3B⁵.

Discussion

Exploration of PGP Patient Reports

The exploration of population genetic data attempted to find patterns in the occurrence of variants and thereby potential subsets within the population that may have certain genetic predispositions. First, “exclusive” genes were found using the largest five unique sets of variants. Here, exclusivity refers to a gene with three or more variants belonging to only one report during pairwise comparison of the five sets. The first and third-largest groups exclusively contained variants on the gene *AGBL4*. The fifth-largest group was the only one to lack variants on *RRM1*. A related exercise to explore in the future is to group reports by conditions, for example, some ethnic groups may have a greater susceptibility for a condition.

The presence of unique sets of variants in the downloaded PGP reports may be partly due to the inclusion of DTC data from multiple sources that have differing technologies. Another possibility is that the presence of unique groups also represent subsets within the population. A majority of unique RSID sets belonged to a few participants. If the larger groups of unique variants are indicative of populations, these smaller sets could be combined if the variant compositions are similar. Potential groupings with the largest unique set was investigated, which encompassed 281 participant reports. One unique set matched, which contained 696 variants and encompassed one participant report. One difficulty for exact groupings is an imperfect match (i.e. Set A matches with Sets B, C, D, and Set E matches with Sets B, C, F). Another is the ambiguity of the commutative property (i.e. Set A matches Set B, but Set B does not necessarily match Set A). It may be interesting to validate these results when given demographic information as labels. Ultimately, the goal of this exercise would be to predict whether an individual may be more likely to have certain variants and therefore increased drug sensitivity based on demographic information.

Future analyses might advance this research further by accounting for the genotypes within participant reports, as the work presented here is based only on the variant identities. Drug sensitivities and conditions are genotype-specific; an individual having a particular RSID with an annotation may possess a genotype that has no significant effect or even a decreased risk of the adverse event. The unique sets found in this study may require additional exploration since the genotypes of the encompassed reports are likely different.

Associated Drugs

Population 23andMe data combined with PharmGKB annotations is a potentially useful resource in pharmacogenetics by indicating related RSIDs and, consequently, drugs. The network of drugs built in this work (Figure 2) was based only on pharmacogenetic and population data. Drug associations may be examined through shared pathways or structural similarities, but a genetic association may encompass those relationships or involve alternative ones. One example is drugs that have a combined therapeutic effect but are chemically different. For example, cyclophosphamide is an alkylating antineoplastic agent with similar affected genes to cisplatin, a platinum-based agent without an alkylating group, and studies have shown encouraging results for combination treatments using the two drugs¹⁸. In a clinical setting, knowledge of drugs associated with genetic profiles may be important when prescribing alternate lines of treatment for an indication. For example, carboplatin and cisplatin are platinum-based chemotherapy drugs that share structured indications on DrugBank; while analogues, the drugs have no shared genes, and therefore one treatment may be prescribed instead of the other to mitigate ADEs.

The modularity of the network was dependent on edge weights between drugs. For two drugs to be strongly correlated using this methodology, the associated variants for both must commonly co-occur in the patient population. While variants on different chromosomes have no association by definition of being unlinked, they may have a nonrandom tendency to be co-inherited, referred to as linkage disequilibrium. One reason for this is a shared function between the variants, which are then associated during selection. Although the typical use for this quantity is for locations on the same chromosome, the original calculation allowed for the consideration of different chromosomes¹⁹. When looking at co-occurring variants affecting weight gain, the analysis demonstrated that variants on different chromosomes may have a nonrandom tendency to co-occur. When the steps were repeated for drugs causing neutropenia, it was observed that some variants on different chromosomes co-occurred while others did not. For example, RSIDs on chromosomes 7 and 12 affected sensitivity to clozapine but did not co-occur in the two largest sets, whereas RSIDs on chromosomes 7 and 13 did for valganciclovir. As previously mentioned, these analyses were performed only for variant identities, and additional examination of individual genotypes is needed.

Successful grouping of drugs, which is assessed in detail below, would indicate that annotated RSIDs are captured in 23andMe data, supporting the potential clinical utility of DTC tests. Table 1 shows that drugs in the same classification tend to share variant annotations and are expected to be grouped in the same modularity class. Measures of precision and recall were calculated by comparing the success of clustering for drugs with expected groupings. Values can be calculated for each drug classification, where each is assigned a modularity class as determined by the Gephi algorithm. The average precision of this methodology was 0.731, and the recall was 0.742. Classes that had fewer than three drugs, among other exceptions, were not included in the averaged values. The green modularity class mentioned in the results lowered these values due to the coverage of two drug classifications that otherwise were clustered with high precision. Two modularity groups (blue and pink) were not included in the calculation and were less successful clusterings from a drug classification perspective; notably, the pink cluster contained five labeled drugs and spanned four categories.

Quantifying precision and recall for this application is difficult because the nature of drug association is uncertain. Here, they were calculated according to drug classification, which is an imperfect measure since drugs with different classifications may share other similarities like drug interactions or chemical structure. For example, iloperidone, an atypical antipsychotic, was clustered with several tricyclic antidepressants due to drug interactions, which were included in the list of “moderate” interactions in the drug’s boxed warning²⁰. Conversely, some drugs of different classifications were grouped in the same modularity class like atenolol and verapamil. While the former is a beta blocker and the latter is a calcium channel blocker, they share one annotated RSID and are both prescribed for angina and high blood pressure.

The levels of precision and recall obtained in this study support the proposed methodology for discovering potential drug associations. Returning to the green cluster from above, fenofibrate, terbinafine, and ticlopidine were grouped drugs that lack clear similarities in structure or indication. It would be interesting to investigate if these unrelated drugs would be removed from the modularity cluster with the introduction of more data or if there is an underlying genetic association that might make an individual prone to ADEs from both drugs. Caffeine was also an associated drug, and it has known effects on psychiatric symptoms and potential interactions with antipsychotic medications^{21,22}. Although this association has been documented in preliminary scientific studies, this is an example of discoveries that may have interesting research or clinical implications.

As mentioned earlier, drug interactions are one possible association depicted by the clustering. One attempt to verify this association was through DrugBank. Paroxetine was chosen for this purpose because its associations were concentrated in two separate clusters. There was no overlap between possibly related drugs through variants and through AEOLUS case reports. The number of drug interactions on DrugBank is rather large, so the list should be curated for significance or another means of verification may be necessary. If drug interactions found through genetic variants can be verified, associations that are unexpected may have clinical and research potential.

Variant-ADE associations in this study focused only on the outcomes for paroxetine, but in the future the network of ADEs should be mapped across a larger set of drugs to obtain a more comprehensive network. Many of the outcomes shown in Figure 2a had one or two connecting edges; the addition of more nodes and edges would facilitate the formation of more distinct clusters. It is interesting that less distinct groups were already formed from this limited demonstration; for example, headache, vomiting, and dizziness commonly occur alongside feelings of nausea. Expanding the network of associated outcomes to include a variety of drugs would further investigate the

notion that ADEs have characteristic pathways that genetic variants may affect. There was no obvious genetic correlation between fatigue, nausea, and suicidal ideation when treated with paroxetine. The variants were located on different chromosomes and affected unrelated genes. An interesting observation is that the variant associated with paroxetine and fatigue affects CYP1A2, and the CYP1A2 enzyme metabolizes caffeine. A recent study demonstrated an increased serum concentration of paroxetine with the coadministration of caffeine¹⁹.

Potential Clinical Utility of DTC-Derived Data

One of the principal aims of this study was to assess the potential clinical utility for DTC genetic testing like 23andMe. These genetic tests have the potential to increase the accessibility of personal genetic data, which can be used as an additional source of information that is more comprehensive and consistent than what is normally provided in patient history forms. 84.1% of annotated variants and 93.5% of drugs were contained within all genetic profiles, which suggests that a majority of variant-drug associations are included in the 23andMe screening. As mentioned previously, the variant-drug associations from PharmGKB were not fully captured in the annotations file due to the limited rule-based approach used in this feasibility study. The categories of some annotations were inconsistent with the findings contained within the summaries. For instance, some associations of toxicity were related to dosage reductions, while an example association within the dosage category involved an increased risk of neutropenia. In particular, an alternate method is necessary to capture these inconsistencies that involve outcomes outside of the curated vocabulary. Obtaining the remaining annotations would increase the number of clinically relevant associations found in 23andMe tests.

While consumer genetic tests are not intended for diagnostic purposes, the information obtained from a population perspective may have clinical utility. Given the results of this study, 23andMe profiles could be used to alert patients and medical practitioners of potential ADEs, prompting a more comprehensive genetic test in the clinic. At the minimum, a screening process could include rapid comparison of a DTC profile like 23andMe to a knowledgebase of RSIDs associated with drug toxicity (e.g., sourced from PharmGKB). The genetic profile for an individual that indicates an increased chance for an adverse event to one drug might also suggest that ADEs are probable for another, whether it is due to drug interactions or a separate affected pathway. The drug and ADE relationships examined in this study demonstrate that valid associations may be obtained by examining genetic information at a population level. That being said, more work is required to validate the findings and to expand the analyses beyond a single drug and drug group.

Limitations

The incorporation of 23andMe profiles into the research and clinical space was one of the principal aims of this study, and the major limitations of this study are correspondingly related to shortcomings of the DTC genetic testing. One of the major shortcomings Lu et al. discussed was the significant genetic variations between different populations. A smaller source for inconsistency is in the self-reporting system of PGP. Additionally, some files were samples of the total report, which leaves the possibility that smaller reports may be sections of profiles rather than being due to genetic ground truth. To support the inclusion of all reports in the analysis, the proportion of the RSIDs contained in each file that were matched to annotated variants were calculated, and the proportions were found to be reasonably similar.

Conclusion

This study demonstrated that associations between drugs can be obtained by examining genetic profiles at a population level. From a research perspective, these associations could guide future research directions or serve as an additional pharmacogenetic resource. Additionally, the survey of RSIDs associated with ADEs suggests that 23andMe genetic tests may have clinical utility in flagging the potential need for full diagnostics.

Acknowledgements

This work was funded in part by National Institutes of Health grants R01LM011364, R01LM011963, and U54GM115677. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Sultana J, Cutroneo P, Trifirò G. Clinical and economic burden of adverse drug reactions. *J Pharmacol Pharmacother.* 2013; 4(Suppl 1): S73-7.
2. Hug BL, Keohane C, Seger DL, Yoon C, Bates DW. The costs of adverse drug events in community hospitals. *Jt Comm J Qual Patient Saf.* 2012; 38(3): 120-6.
3. Questions and Answers on FDA's Adverse Event Reporting System (FAERS) [Internet]. US Food and Drug Administration. 2017 [cited 29 August 2017]. Available from: <https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm>
4. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. *Sci Data.* 2016; 3: 160026.
5. M. Whirl-Carrillo, E.M. McDonagh, J. M. Hebert, L. Gong, K. Sangkuhl, C.F. Thorn, R.B. Altman and T.E. Klein. "Pharmacogenomics Knowledge for Personalized Medicine" *Clinical Pharmacology & Therapeutics* (2012) 92(4): 414-417.
6. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001 Jan 1; 29(1): 308-11.
7. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 2014 Jan 1; 42(1): D1091-7.
8. Elliott LS, Henderson JC, Neradilek MB, Moyer NA, Ashcraft KC, Thirumaran RK. Clinical impact of pharmacogenetic profiling with a clinical decision support tool in polypharmacy home health patients: A prospective pilot randomized controlled trial. *Plos One.* 2017; 12(2).
9. Olson MC, Maciel A, Garipey JF, Cullors A, Saldivar JS, Taylor D, et al. Clinical impact of pharmacogenetic-guided treatment for patients exhibiting neuropsychiatric Disorders: a randomized controlled trial. *Prim Care Companion CNS Disord.* 2017; 19(2): 16m02036.
10. Giacomini KM, Yee SW, Mushiroda T, Weinshilboum RM, Ratain MJ, Kubo M. Genome-wide association studies of drug response and toxicity: an opportunity for genome medicine. *Nat Rev Drug Discov.* 2017;16(1):1.
11. Lu M, Lewis CM, Traylor M. Pharmacogenetic testing through the direct-to-consumer genetic testing company 23andMe. *BMC Med Genomics.* 2017; 10(1): 47.
12. Ball M, Bobe J, Chou M, Clegg T, Estep P, Lunshof J et al. Harvard Personal Genome Project: lessons from participatory public research. *Genome Medicine.* 2014; 6(2): 10.
13. Mental Health Disorders [Internet]. Drugs.com. 2017 [cited 26 September 2017]. Available from: <https://www.drugs.com/mental-health.html>
14. Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media.*
15. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks.
16. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. PubChem Substance and Compound databases. *Nucleic Acids Res.* 2016 Jan 4; 44(D1): D1202-13.
17. Iloperidone - Drug Summary [Internet]. Prescriber's Digital Reference. 2017 [cited 28 September 2017]. Available from: <http://www.pdr.net/drug-summary/Fanapt-iloperidone-429#topPage>
18. Dasari S, and Tchounwou PB. Cisplatin in cancer therapy: molecular mechanisms of action. *Eur J Pharmacol.* 2014 Oct 5; 0: 364-378.
19. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat Rev Genet.* 2008 Jun; 9(6): 477–485.
20. Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353-D361 (2017).
21. Szopa A, Poleszak E, Wyska E, Serefko A, Wosko S, Wlaza A, et al. Caffeine enhances the antidepressant-like activity of common antidepressant drugs in the forced swim test in mice. *Naunyn-Schmiedeberg's Arch Pharmacol.* 2016; 389: 211–221.
22. Broderick PJ, Benjamin AB, Dennis LW. Caffeine and psychiatric medication interactions: a review. *J Okla State Med Assoc.* 2005 Aug; 98(8): 380-4.